Microbial network disturbances in relapsing refractory Crohn's disease

Bahtiyar Yilmaz[1,2,10], Pascal Juillerat[1,2,10], Ove Øyås[3], Charlotte Ramon[3], Francisco Damian Bravo[2], Yannick Franc[4], Nicolas Fournier[4], Pierre Michetti[5], Christoph Mueller[6], Markus Geuking[1], Valerie E.H. Pittet[4], Michel H. Maillard[7], Gerhard Rogler[8], Swiss IBD Cohort Investigators[9], Reiner Wiest[1,2], Jörg Stelling[3] & Andrew J. Macpherson[1,2*]

[1]Maurice Müller Laboratories, Department for Biomedical Research, University of Bern, Bern, Switzerland. [2]Department of Visceral Surgery and Medicine, Bern University Hospital, University of Bern, Bern, Switzerland. [3]Department of Biosystems Science and Engineering and SIB Swiss Institute of Bioinformatics, ETH Zurich, Basel, Switzerland. [4]Institute of Social and Preventive Medicine (IUMSP), Lausanne University Hospital, Lausanne, Switzerland. [5]Gastroenterology La Source-Beaulieu, Lausanne, Switzerland; Division of Gastroenterology, Lausanne University Hospital, Lausanne, Switzerland. [6]Division of Experimental Pathology, Institute of Pathology, University of Bern, Bern, Switzerland. [7]Service of Gastroenterology and Hepatology, Department of Medicine, Centre Hospitalier Universitaire Vaudois and University of Lausanne, Lausanne, Switzerland. [8]Clinic for Gastroenterology and Hepatology, University Hospital Zurich, University of Zurich, Zurich, Switzerland. [9]A full list of members and affiliations appears at the end of the paper. [10]These authors contributed equally to this work.

[*]**Corresponding author**: andrew.macpherson@dbmr.unibe.ch

University Hospital Bern (Inselspital), D117, BHH, Bern 3010, Switzerland, Tel. +41 31 632 8025

ABSTRACT

Inflammatory bowel diseases (IBD) can be broadly divided into Crohn's disease (CD) and ulcerative colitis (UC) from their clinical phenotypes. Over 150 host susceptibility genes have been described, although most overlap between CD, UC and their subtypes, and they do not adequately account for the overall incidence or the highly variable severity of disease. Replicating key findings between two long-term IBD cohorts we have defined distinct networks of taxa associations within intestinal biopsies of CD and UC patients. Disturbances in an association network containing taxa of the Lachnospiraceae and Ruminococcaceae families, typically producing short chain fatty acids, characterize frequently relapsing disease and poor responses to treatment with anti-TNF-α therapeutic antibodies. Alterations of taxa within this network also characterize risk of later disease recurrence of patients in remission after the active inflamed segment of CD has been surgically removed.

INTRODUCTION

There is strong evidence indicating that the intestinal microbiota is responsible for triggering inflammatory bowel disease (IBD): animal models of chronic intestinal inflammation are almost all mitigated when re-derived to germ-free status[1,2]; humans with IBD have altered reactivity to their intestinal microbes[3]; the genetic loci for human susceptibility to Crohn's disease (CD) and/or ulcerative colitis (UC) mainly encode proteins required for innate or adaptive mucosal immune defense mechanisms and overlap with genes known to cause microbiota-sensitive intestinal inflammation in animal models[4]. Whilst no persistent pathogen has been consistently identified as being responsible for IBD, strategies reducing microbiota exposure such as surgical fecal stream diversion[5], elemental diets or antibiotic treatment can have beneficial effects in particular IBD cases[6-8].

The intestinal microbiota stabilizes during childhood, but with inter-individual variations[9-11] and fluctuations according to diet[12] or intestinal transit time[13,14]. Microbial taxa compositions in IBD patients have been extensively studied (Extended Data Fig. 1a). These different studies have been carried out with different microbiological and sequencing techniques, on various ethnic populations, at different ages and stages in the trajectory of disease, using fecal or tissue specimens. Alterations in the relative proportions of taxa within the Bacteroidetes, Firmicutes and Proteobacteria in CD patients compared with UC or non-IBD subjects are generally reported. Despite this, the overall picture remains very inconsistent (Extended Data Fig. 1 and 2).

Microbiota differences may account for i) a gap in the IBD prevalence that is inadequately explained by known CD/UC genetic susceptibilities[15-17]; ii) phenotypic differences in the severity and distribution of disease that are only partly explained from genetic and immunological studies[15,18-20]; and iii) a significant proportion of patients that are poorly responsive to current selective or non-selective immunomodulatory treatments[21-23].

Within CD and UC there is immense variability in disease distribution, severity and responsiveness to treatment. These clinical disease behaviors are poorly predicted by host genetic susceptibility markers. Disease behavior, treatment responsiveness and environmental influences have not been comprehensively considered in studies of microbiota patterns in IBD so far (Extended Data Fig. 1b). Our hypothesis here was

that microbiota composition and taxa networks underlie the phenotypic variability of established CD and UC that could be addressed with robust deep longitudinal information over many years for each patient. Given that important and potentially correctable differences should be reproducible in the same (central European) environment, we have determined microbiota changes according to disease phenotype that are robust to replication between two independent long-term longitudinal cohorts of patients. We show here microbiota networks that are reproducibly related to long-term disease type and severity, the extent and position of the involved intestinal segments and responsiveness to different treatments.

RESULTS

Distinct microbial communities according to CD or UC diagnosis in IBD patients compared with controls

The microbial communities of IBD patients were assessed using 16S sequencing in biopsy samples obtained from a tissue biopsy biobank of the Swiss IBD cohort (*Cohort 1*). To validate the results a second independent cohort of patient biopsies (*Cohort 2*) including healthy, non-IBD controls was collected from Bern gastroenterology clinics (Supplementary Table 1).

Matching clinical deep metadata sets were assembled prospectively and retrospectively from patient records for both cohorts. In addition to gender, age, marital status, family history, specifics of diagnosis, its timing, and disease subtype at the outset, contemporaneously collected follow-up information for each patient included into trajectories of disease evolution with blood parameters, environmental exposures (alcohol and tobacco use), daily exercise, BMI, medication time-courses (antibiotics, corticosteroids, mesalamine, and immunosuppressants), nutritional supplements, disease complications, surgical resections and inter-current comorbidities.

As expected from known taxa predilection in the human large intestine[24], the dominant bacterial phylotypes in the gastrointestinal (GI) tract of IBD patients from both cohorts and healthy controls were Bacteroidetes, Firmicutes and Proteobacteria, with a smaller proportions of Fusobacteria, Actinobacteria and Tenericutes (Supplementary Tables 1 and 2). Average Firmicutes proportions are known to be higher in comparison with Bacteroidetes and we observed this in all GI segments of UC compared with CD in both cohorts but with very wide variability across subjects and sites (Extended Data Fig. 3a, b). As also expected from previous studies[25], overall alpha diversity measured according to the Shannon and Simpson indices demonstrated significantly reduced diversity in CD samples compared with UC samples in both patient cohorts, and compared with healthy, non-IBD controls in *Cohort 2* (Extended Data Fig. 3c,d).

The distinctiveness of the CD microbiota was confirmed by beta diversity analysis, demonstrating a clustering of samples according to IBD diagnosis for both cohorts with the same homogeneity by Bray-Curtis dissimilarity metrics ($p<0.001$) (Fig. 1a,b) as well as UniFrac distance metrics ($p<0.001$) (Extended Data Fig. 3e-h).

We next evaluated specific microbial features associated with disease phenotypes that could be replicated at genus level in both independent cohorts (Fig. 1c). Comparison of relative abundances in CD and UC patients revealed multiple concordant significant differences at phylum (Actinobacteria, Firmicutes, Bacteroidetes and Proteobacteria) and genus level (*Bifidobacterium, Coriobacteriaceae•, Barnesiellaceae•••,* Clostridiales genus••, *Prevotella* (Prevotellaceae family)*, Prevotella•••,*

(*Paraprevotellaceae* family)*, Lachnospira, Faecalibacterium, Coprococcus, Roseburia,* two *Lachnospiraceae genera* (• and ••)*,* two *Ruminococcaceae genera* (• and ••)*, Ruminococcus, Enterobacteriaceae•,* and *Streptococcaceae••* (Fig. 1c and Extended Data Fig. 3i,j and Supplementary Table 3).

From the identification of taxa that are selectively associated with CD, UC or controls that are robust to replication in two independent cohorts from central European environment, we compared these results compared with an unsupervised meta-analysis of previously published studies carried out in different contexts (Extended Data Fig. 1a). In general, different genera within Firmicutes are reported to be reduced in CD compared with controls, or IBD (where the study concerned grouped CD+UC) compared with controls (Group 1 in Extended Data Fig. 1a); further genera within Proteobacteria are increased in CD or CD+UC compared with controls (Group 2 in Extended Data Fig. 1a). For other taxa, our meta-analysis showed results that were largely inconsistent between different studies (Group 3 in Extended Data Fig. 1a) regardless of age, gender, ethnicity, sample type or sequencing method. We now show in Extended Data Fig. 1a (bold taxa labels) that our results verify the previous data for 8 of 11 taxa in Group 1 and 3 of 4 taxa in Group 2: Our results also show 18 significant additional taxa differences between CD and controls, or UC and controls that are robust to replication between the two cohorts in the heterogeneous Group 3. These changes that are seen between CD and UC are largely a consequence of differences between the CD microbiota and non-IBD controls, whereas UC microbial consortia are relatively similar to healthy subjects (Extended Data Fig. 4a,b).

Using our extended portfolio of emblematic taxa, machine learning enabled prediction of CD and UC clusters with 87.7% and 83.3% prediction accuracy for *Cohort 1* and *Cohort 2*, respectively (Supplementary Table 4). Focusing on multiple predicted OTUs within specific genera, we confirmed that *Ruminococcus gnavus* was consistently increased in abundance in CD compared to UC, and *Faecalibacterium prausnitzii* was consistently reduced in both cohorts[26,27] (Extended Data Fig. 4c,d).

We next used genome-scale metabolic models to analyze our sequencing data and elucidate how metabolism is modulated within microbial IBD communities[28]. Principal component analysis (PCA) analysis illustrates robust data at the metabolic reaction level with consistent metabolic subsystems increased in CD belonging to B-vitamin and LPS biosynthesis, heparan sulfate and chondroitin sulfate degradation, and fatty acid oxidation (Extended Data Fig. 4e-j).

We also carried out a meta-analysis comparing IBD animal models including dogs, cats, and laboratory mice with the available human studies. In an unsupervised analysis, the microbiota clustered according to species (Extended Data Fig. 5a), and genetic models were the most representative of the changes seen in human IBD patients (Extended Data Fig. 5b,c).

In summary, the distinct CD and UC disease groups are characterized by altered bacterial composition and lower diversity in CD patients compared to UC patients and non-IBD subjects (Fig. 1a-c, Extended Data Fig. 3, 4a,b). Bacterial richness and diversity between UC and non-IBD were less pronounced (Extended Data Fig. 3c-f). Our results indicated that there is a greater net loss of potentially beneficial taxa in CD compared with UC (Fig. 1c, Extended Data Fig. 3i,j and 4a,b). These included *Faecalibacterium* capable of butyrate production*;* the secondary bile acid producer *Oscillospira*; the dietary carbohydrate utilizer *Bifidobacterium*; *Ruminococcus* which degrades mucus; and the complex carbohydrate utilizer *Blautia*[29,30].

Longitudinal comparisons of microbial community compositions between CD and UC patients

Most of the known IBD genetic risk factors that are shared between CD and UC are uninformative about disease subtypes[20,31]. Within distinctive clinical pathology and sites of GI tract involvement that define CD or UC[32], disease subtypes may shape the severity and treatment responsiveness for individual patients. Whilst IBD is characterized by chronic relapsing-remitting intestinal inflammation, current published studies have characterized these diseases largely in generic terms that might be insensitive to the longitudinal disease course and/or treatment responsiveness (Extended Data Fig. 1b). Our study was therefore designed to combine deep longitudinal clinical phenotypes between CD and UC and the different disease subtypes, allowing us to replicate key findings using the two independent longitudinal cohorts.

In a subset of our patients, we had biopsies that were taken at different times in the clinical course of disease. In both cohorts, ordination analysis showed that microbiota profiles were personalized according to each individual, over periods ranging up to 9 years (Extended Data Fig. 6a). Individual taxa were generally stable over time (Extended Data Fig. 6b) even when there had been an inter-current change in disease activity defined according to clinical phenotyping (Extended Data Fig. 6c).

To address the question of variation in the microbiota compositions according to parameters of disease activity by alternative methods, we assessed the relationship of the microbiota composition to disease activity according to clinical indices of activity (CDAI for CD; MTWAI for UC) or fecal calprotectin measurements. Of 83 taxa analyzed, only *Enterobacteriaceae*• and *Klebsiella* in CD (Extended Data Fig. 6d), *Ruminococcus*••• and *Prevotella* in UC (Extended Data Fig. 6e) showed consistent compositional changes between the cohorts that were significantly aligned with clinical assessments of the disease activity (Fig. 1d,e), and according to the calprotectin biomarker (Extended Data Fig. 6d,e).

From these findings that microbiota composition profiles were mainly personalized according to the patient rather than the current disease activity, we proceeded to consider whether microbiotas within individuals could be related to their overall disease course or treatment responsiveness.

Co-occurrence pattern analysis of the intestinal ecosystems of IBD patients

We next used co-occurrence analysis to examine potential relationships between different taxa in either CD or UC patients. The complex microbial ecosystem of the human intestinal tract is unevenly influenced by individual taxa within different microbial consortia[10,11,24]: the ecological roles of individual community members are fundamentally important for overall biodiversity and key to human health[33,34]. Computationally derived co-occurrence associations are useful to infer influences between taxa in a range of ecosystems and are a starting point to examine the mechanisms of community construction and maintenance[35,36].

We first tested whether co-occurrence associations would differ strongly across different segments of the GI tract. All patients in both cohorts of our study had received prior intestinal purging to remove faecal material and allow safe endoscopy with biopsies. Since purging involves substantial fluid flow along the length of the small and large intestine, as expected we did not see major microbial profile differences between different intestinal segments of individual patients (Extended Data Fig. 7a-f, Supplementary Table 5). Likely also because of the mixing effect of purging, we also found that the mucosal microbiota diversity

of inflamed and non-inflamed segments within the same patient were indistinguishable, independently of disease type (Extended Data Fig.7g-j, Supplementary Table 5).

We then used all sequenced biopsies from different locations of the IBD patients' gut as replicate sampling groups for CD and UC independent ecosystems to address whether co-occurrence patterns among the microbial communities are the same in the different pathological contexts. We identified 2 consistent modules of co-occurring gut microbes for CD (CD_A and CD_B) and for UC (UC_A and UC_B) (Fig. 2 and Extended Data Fig. 8).

We found similar complexity and composition of the networks for both disease groups in each of the two cohorts [Supplementary Table 6 and 7]: The highest 'betweenness centrality scores' conceivably identifying keystone species within each ecosystem, showed that Firmicutes was a potential keystone phylum determining the network in both CD and UC disease groups of both cohorts (Extended Data Fig. 8a,b). However, Tenericutes for CD in *Cohort 1* and Fusobacteria for UC in both cohorts were also likely important keystone phyla (Extended Data Fig. 8a,b). Measures of eigenvector and betweenness centrality showed in both cohorts that nodes contribute equivalently to the networks although there were more complex relationships of microbial taxa in CD (Extended Data Fig. 8c,d).

Based on betweenness centrality of each vertex, the top genera identified as keystone taxa in both cohorts with higher scores in CD compared to UC were *Streptococcus, Ruminococcus, Bifidobacterium, Propionibacterium, Corynebacterium, Coprococcus, and Blautia* (Extended Data Fig. 8e,f). Further, the degree centrality measure helps identify the most prominent and influential taxa that we might miss with betweenness and Eigenvector centrality measures. Interestingly, *Faecalibacterium, Ruminococcus•••* of the *Lachnospiraceae* family, *Fusobacterium, Haemophilus*, and *Parabacteroides* have some of the lowest betweenness scores although they show amongst the highest out-degree scores (number of edges originating in a node) in both disease groups of both cohorts, implying that *Faecalibacterium* and *Ruminococcus•••* are some of the most influential taxa in *Cohort 1* (Extended Data Fig. 8g,h) and *Cohort 2* (Extended Data Fig. 8i,j) independently of disease status. Overall, whilst there is similar complexity of the networks for both disease groups with relatively consistent modules of co-occurring microbial taxa within each ecosystem, the module labelled Cluster CD_A in both cohorts connects nodes of *Lachnospira, Blautia, Dorea, Coprococcus, Ruminococcus, Ruminococcus•••*, *Faecalibacterium, Roseburia, Oscillospira and Bilophila* (Fig. 2): a similar Cluster UC_B is present in patients with UC, although it is less connected, and overlaps with a Cluster $Cont_B$ that is found in non-IBD controls (Fig. 2). As shown subsequently in the paper, reductions in members of Cluster CD_A within the microbiota of CD patients generally characterize a worse outcome, less healthy lifestyles, poor responses to treatment and are associated with an increased risk of subsequent relapse in patients where the active segment of CD has been surgically removed.

Identification of critical variables in shaping the intestinal microbiota of IBD patients

Analyses of variability of the microbiota with chronic disease phenotype parameters suffer from the limitations that i) correlations seen may be direct or indirect and ii) that the strength of these correlations may be disproportionate if taken in isolation from parameters of subject ageing, weight or other environmental influences. To allow a reasonable perspective of the relative influences of disease

parameters or other personal variables, we then used the algorithms in the Hierarchical All-against-All Association (HAllA) tool for multi-resolution associations in high-dimensional, heterogeneous datasets with high power [37]. Analysis performed at phylum (Fig. 3a,c) and genus rank (Fig. 3b,d) identified BMI, age at sampling/diagnosis as the most important variables correlating with gut microbiota of IBD patients. To a lesser extent, disease type, location and behavior, and prior surgery were also important variables showing interactions with more than 60-70% of most abundant taxa of IBD patients. Overall life styles (including sport, smoking and alcohol consumption) were correlated with the gut microbiota profile in *Cohort 1* (Fig. 3a-d): of these, smoking was the most significant factor. In contrast, biopsy location and inflammation status at the time of sampling were correlated with only a few taxa, which were largely inconsistent between cohorts.

Correlations with patient age and BMI in HAllA are consistent with published data showing strong effects of nutrition and intestinal physiological parameters on the microbiota[12-14]. Although both CD and UC show age-dependent incidence, and active disease may cause weight loss, the greater strength of microbiota association with age at diagnosis than disease type or BMI compared with disease activity argues that these are likely not dependent variables. Taken together with the similarities in taxa co-occurrence analysis at phylum and genus level between patients with CD or UC (Fig. 2), the relatively weaker correlations of the microbiota with disease parameters and responsiveness to treatment (Fig. 3) indicate that detailed minor compositional changes may be associated with disease type or subtype on a background of larger effects of body habitus or age.

The assessment of critical variables that may be related to the microbial profile including disease status, therapeutics and environmental factors

We next focused specifically on the relationship of the microbiota within the grouping of either CD or UC patients, according to the detailed disease phenotype or subtype, known trigger factors and treatment responsiveness using a sparse multivariate statistical approach.

We first analysed whether the patient's lifestyle such as alcohol consumption, smoking habit, or physical activity were associated with alterations of the gut microbiota composition (Supplementary Table 8). We only identified significant changes in microbial taxonomy (in every case corrected for multiple testing) in relation to host lifestyle in CD patients. Increasing exercise was positively correlated with the relative abundance of taxa from the Firmicutes and Bacteroidetes phyla, including *Clostridales*•, *Blautia*, *Barnesiellaceae*•, *Faecalibacterium*, *Eryselotrichaceae*•, *Ruminococcaceae*• genera and negatively correlated with *Veillonella* abundance (Extended Data Fig. 9a). Tobacco smoking or alcohol consumption *Parabacteroides* and *Clostridales*•• were respectively associated with in CD: both were negatively associated with *Sutterella* (Extended Data Fig. 9b,c). *Parabacteroides* was reduced in familial CD patients (Extended Data Fig. 9d). Notably both exercise in CD patients which is indicative of a healthier life-style [23] is associated with taxa including *Clostridiales*••, *Blautia*, *Faecalibacterium*, *Eryselotrichaceae*•, *Ruminococcaceae*• that are represented in Cluster CD_A of the co-occurrence analysis (Fig. 2), although we cannot distinguish whether the alterations are either bias due to reduced disease activity in those able to participate in sport[38] or a beneficial alteration in microbiota composition.

Disease therapy including monoclonal antibody treatment against TNF-α and steroid use are effective treatments for many IBD patients[39-41]. Unfortunately, not all patients are responsive to these therapies likely because of the variable redundancy of different triggering mechanisms of intestinal inflammation and the variable redundancy of TNF-α as an inflammatory cytokine[42]. We asked whether the responsiveness of IBD patients to different therapies could be correlated with differences in microbiota composition. Initial microbial analysis of CD and UC samples showed that there was no significant difference in alpha diversity for either cohort based on whether the patient had responded to treatment (Extended Data Fig. 9e,f). However, we did find significant differences in gut microbial composition between patients responding or not responding to TNF-α inhibitor therapy (Fig. 4a-c). Although there were no significant differences in the small numbers of UC patients that had been treated in both cohorts, comparison of the CD groups responding to treatment versus the CD patients failing to respond to treatment, showed and replicated increased *Bifidobacterium*, *Collinsella*, *Lachnospira, Lachnospiraceae•, Roseburia, Eggerthella* taxa and reduced *Phascolarctobacterium* associated with treatment success (Fig. 4b,c). *Bifidobacterium* are well recognised for their capacity for metabolism of oligosaccharides that may be resistant to host digestion and have been used as probiotics in IBD[43]. *Lachnospiraceae•* and *Roseburia* are both members of association CD_A (Fig. 2) and are characterized by production of short chain fatty acids that are a carbon source for intestinal epithelial cells and inducers of regulatory T cells[44,45]. *Phascolarctobacterium* is common in the human GI tract[46], and can produce acetate and propionate, although it has also been associated with inflammatory dysbiosis in animal models[47].

We asked whether the responsiveness of IBD patients to corticosteroid therapies could also be correlated with differences in microbiota composition. However, although significant differences in gut microbial taxa composition between groups responding to treatment versus groups failing to respond were observed in the PCoA analysis of both cohorts (Extended Data Fig. 9i,j), unlike the results with anti-TNF-α biological treatments, these were not replicated between the different cohorts (Fig. 4d, Extended Data Fig.9k-m).

A major burden for patients with CD and UC is that the diseases follow a relapsing-remitting course of recurrent exacerbations and symptomatic improvements over many years, with unpredictable effects on morbidity, quality of life and the need for healthcare (Supplementary Table 9). The analysis of samples from patients with a quiescent or a relapsing course showed that whilst there were no significant differences in microbial alpha diversities either cohort, in CD, *Eggerthella, Clostridiales••,* and *Oscillospira* showed consistent replicated increases in relative abundance in patients with quiescent disease over time while *Enterobacteriaceae•, Enterobacteriaceae••* and *Klebsiella* were associated with a more severe clinical course (Fig. 1d,e, Extended Data Fig. 9n-r). *Oscillospira* associated with the more benign course of CD are also members of association Cluster CD_A in *Cohort 1* and in Cluster CD_C together with *Eggerthella* in *Cohort 2* (Fig. 2).

Disease location is secondary determinant in gut microbiota profile of IBD patients

Separating the CD patients according to whether they had colonic disease, or ileal CD disease (comprising ileal CD and ileo-colonic CD subtypes combined), we found no significant differences between non-IBD, colonic UC, and colonic CD samples, although the gut microbiota composition of patients with colonic CD

was closer to the microbiota of patients with UC than to that of patients with ileal CD, as previously reported [32] (Fig. 5a-c). Metabolic reaction set analysis of IBD subtype microbiotas, that were consistently seen across the two patient cohorts, revealed that LPS biosynthesis, vitamin B2 and vitamin B7 metabolism were higher in colonic CD subtype compared with (colonic) UC (Fig. 5d), and LPS biosynthesis, peptide and folate metabolism, fatty acid synthesis, chondroitin sulphate and heparan sulphate degradation were functionally elevated in patients with ileal CD subtype compared with UC (Fig. 5e). Glycerophospolipid metabolism and cell wall biosynthesis were increased in UC microbiota compared with colonic CD or ileal CD respectively. There was no significant metabolic reaction set differences between ileal and colonic CD, although any distinction may be confounded by colonic purgatives prior to sampling as discussed previously.

Post-surgical CD patients have altered intestinal microbiota prior to disease recurrence

Although the outlook for IBD patients has been transformed by the advent of treatments that can be directed against selected pro-inflammatory pathways, surgery is still required in up to 40% of CD and 30% of UC patients at some point during their lifetime[48,49]. UC surgery normally comprises removal of the entire colon and hence the source of the intestinal inflammation. Before the advent of current biological therapies, clinical studies established that removal of inflamed intestinal segments in CD cannot cure the disease, because without ongoing treatment CD recurs in over 70% of patients at the anastomotic margin where previously uninflamed segments have been joined [50]. Post-surgical patients represent and extremely interesting group because they usually do not have detectable Crohn's disease (since it has been surgically removed), but they will almost inevitably acquire it again unless they receive immunosuppressive medications. Thus, the microbiota in patients after surgery may be emblematic of the patient susceptible to CD (potential CD), without it being present at the time of sampling.

We therefore compared the post-operative microbiota in CD patients that had already undergone resection of the ileo-cecal region or a segment or small intestine, with CD patients that had not needed an operation. The operated CD patients had distinct gut microbial profiles with reduced species richness in CD samples from both cohorts collected >3 months post-operatively (Fig. 6a,b). Further, composition differences of the intestinal microbiota were found between inactive CD patients with or without prior surgery (Fig. 6c, d). The taxa that were principally included replicated reductions in the following members of Cluster CD_A, *Ruminococcaceae*•, *Oscillospira*, *Ruminococcus* and *Bilophila*. In both cohorts *Parabacteroides* and *Clostridiales*• were also reduced in inactive post-surgical patients, and *Enterobacteriaceae*•• were significantly and reproducibly increased (Fig. 6e,f). These compositional changes in post-surgical CD patients without active disease, but with an almost inevitable probability of relapse in a previously uninvolved intestinal segment if untreated (biased towards reductions in Cluster CD_A) may be capable of triggering CD relapse.

The implication that a disrupted composition of Cluster CD_A makes patients susceptible to CD relapse, even in post-surgical CD patients who temporarily have no activity because the active segment has been removed, implies that this cluster may be a biomarker of resistance to relapse of active disease independently of current inflammatory activity. It would follow that Cluster CD_A taxa are not themselves intrinsically associated with CD disease activity, unlike the *Enterobacteriaceae*• (Extended Data Fig. 6d).

We found no significant changes in the Cluster CD_A taxa in longitudinally studied patients judged by alterations in disease severity status over time (Extended Data Fig. 10a). As an alternative approach, we also asked whether the calprotectin level could be a more sensitive luminal inflammatory biomarker for possible alterations in the association cluster CD_A with increased intestinal inflammation. Again, no CD_A taxon showed a significant correlation with calprotectin levels (Extended Data Fig. 10b). Overall, this data supports the idea of importance to this cluster of taxa to inhibit relapse, rather than being associated as a biomarker for the patients with the most inflammatory and immunologically active disease.

DISCUSSION

In this paper, we report the relationship of intestinal biopsy microbiotas from patients with UC and CD taken from two cohorts of patients over many years and capturing information about clinical phenotypes including responsiveness to different treatments, life-style differences and the severity of the relapsing-remitting clinical course. In our patient cohorts relatively few taxa were shown to be correlated with the current activity of disease, whether assessed by clinical phenotyping or by the fecal calprotectin inflammatory biomarker.

We have confirmed previous reports of the microbiota differences between CD patients and controls, and UC patients and controls. In addition to the known associations of *Faecalibacterium* and the Firmicutes phylum with UC, and Bacteroidetes and *Ruminococcus*••• with CD, we have now shown (and replicated between cohorts) that other Firmicutes genera (including *Lachnospiraceae•, Clostridia••, Roseburia, Blautia, Ruminococcus* genus of the *Ruminococcaceae* family) are reduced in CD compared with UC, whilst a *Ruminococcus*••• genus of the *Lachnospiraceae* family was increased in CD compared with UC.

That the dysbiosis we observed may be causal is supported by the special situation in patients where CD segments have been surgically removed, yet relapse in another previously uninvolved part of the intestine is extremely likely without immunosuppression.

In line with other studies that show major effects from nutrition and host intestinal physiology on the microbiota[13,14], we have found that the differences that are attributable to CD or UC, or to the different lifestyle or therapeutic factors are far less than those engendered by age or body habitus. Since IBD are highly heterogeneous diseases, we need to be cautious about associations between microbial gut profiles and clinical phenotypes with small but significant effect sizes. Nevertheless, a major biomedical problem is to understand why the phenotypic behavior and therapeutic responsiveness of CD and UC and its subtypes is so different between patients. Here we could show and replicate between cohorts that increased *Bifidobacterium*, *Collinsella*, *Lachnospira, Lachnospiraceae•, Roseburia, Eggerthella* taxa were related to the successful outcome of anti-TNF-α treatment in CD. This is consistent with the finding that *Lachnospiraceae•* and *Roseburia* taxa are reduced in CD, and that these taxa produce short chain fatty acids as a carbon source for intestinal epithelial cells and induction of regulatory T cells[44,45]. We found that *Roseburia* was a likely keystone taxon in microbial consortial association analysis of CD but not UC. In murine and *in vitro* models *Roseburia* can promote innate gut immunity of antimicrobial peptide up-regulation and improved gut barrier function in addition to regulatory T cell induction[51]. Interestingly, reduction of two *Lachnospiraceae*, *Blautia* and *Roseburia* also predict pouchitis in patients that have undergone colectomy with reconstruction of a pouch of small intestine to substitute the rectum[52]. Whilst

reduced *Roseburia* has also been reported compared with controls in UC[53], our findings are that the effect of the taxon on the clinical course is strongest in CD. We also found that *Clostridiales*•, *Blautia*, *Faecalibacterium*, *Eryselotrichaceae*•, *Ruminococcaceae*• were associated with taking regular exercise in CD. *Blautia*, *Faecalibacterium* and *Ruminococcus* are all likely keystone taxa in CD and UC, but with greater connectivity in the association analysis in CD. Many of the above taxa are connected in Cluster CD_A in the association analysis (Fig. 2).

Many of the taxa that are altered in association with a poor prognosis of disease course, therapeutic unresponsiveness, a less healthy lifestyle or likelihood of relapse in postsurgical patients are within the Cluster CD_A, identified through taxa association analysis. This contains taxa characterized by short chain fatty acid production through different metabolic pathways[54] which may increase the availability of intraluminal intestinal oxygen and allow facultative aerobes such as *Enterobacteriaceae* family to bloom through alterations in intestinal epithelial cell metabolism[55]. We found no significant correlations between intestinal inflammation and relative abundance of CD_A taxa measured either according to clinical assessment of alterations in disease activity or in relation to the fecal calprotectin inflammatory biomarker. Taken together, this supports the idea of importance to this cluster of taxa to inhibit the disease, likely in a particular context of individual patient genetic susceptibility, rather than merely being associated as a biomarker for the patients with the most inflammatory and immunologically active disease.

Our data have shown a cluster of taxa including those characterized by their short chain fatty acid production that generally distinguish CD from UC, and are predictive of a quiescent course, a healthier lifestyle and responsiveness to anti-TNF-α therapy. These taxa have been replicated across two patient cohorts in the study and are deficient in patients after surgical resection that can be considered at very high risk for CD relapse despite removal of the earlier diseased intestinal segments. There is considerable interest in manipulation of the microbiota as an adjunctive therapy to improve the course and responsiveness of disease. Our results predict that fecal transplants or directed manipulation of the intestinal microbiota may be most successful where defects in Cluster CD_A can be corrected.

AUTHOR CONTRIBUTIONS

A.J.M. conceived, designed, and supervised the study. B.Y. performed all the experiments, analyzed the data, and wrote the manuscript with A.J.M. P.J., organized and collected the samples of the second cohort. P.J. F.D.B., Y.F., N.F. and M.G. were involved in data curation. O.O., C.R. carried out metabolic reaction analysis and J.S. supervised the analysis. A.J.M., P.J. P.M., C.M., G.R., R.W. and Swiss IBD cohort investigators acquired patient samples and detailed structured clinical phenotypes.

ACKNOWLEDGMENTS

We thank all patients and the members of the Swiss IBD cohort and Bern cohort for their commitment. We also thank the staff of the University Hospital of Visceral Medicine and Surgery, and the Bern City Hospitals led by Frank Seibold and Radu Tutuian for obtaining samples in *Cohort 2*. This research was supported by Systems X (GutX) to A.J.M. and J.S., the Swiss IBD cohort (Grant 33CS30-148422) to G.R., A.J.M and C.M. The founding institutions had no role in study design, analysis, or interpretation of the

results. We thank Dr. Gholamali Rahnavard and Dr. Curtis Huttenhower (Department of Biostatistics, Harvard T.H. Chan School of Public Health, USA) for their help in using HAllA pipeline. We also thank to Jessica Harrell Rieder, Alexandra Suter, Sarah Brand, Catherine Mooser, William Kwong Chung, and Judith Hugenschmidt for helping to B.Y. during the process of sample preparation. We also thank Dr. George Weingart (Department of Biostatistics, Harvard T.H. Chan School of Public Health, USA) for his enormous help on optimization MaAsLin running in MacOS platform using R.

Competing Interests Statement The authors declare no competing financial interests.

Other contributing authors are as follows: Swiss IBD Cohort Investigators: Karim Abdelrahman, Gentiana Ademi, Patrick Aepli, Amman Thomas, Claudia Anderegg, Anca-Teodora Antonino, Eva Archanioti, Eviano Arrigoni, Diana Bakker de Jong, Bruno Balsiger, Polat Bastürk, Peter Bauerfeind, Andrea Becocci, Dominique Belli, José M. Bengoa, Luc Biedermann, Janek Binek, Mirjam Blattmann, Stephan Boehm, Tujana Boldanova, Jan Borovicka, Christian P. Braegger, Stephan Brand, Lukas Brügger, Simon Brunner, Patrick Bühr, Bernard Burnand, Sabine Burk, Emanuel Burri, Sophie Buyse, Dahlia-Thao Cao, Ove Carstens, Dominique H. Criblez, Sophie Cunningham, Fabrizia D'Angelo, Philippe de Saussure, Lukas Degen, Joakim Delarive, Christopher Doerig, Barbara Dora, Susan Drerup, Mara Egger, Ali El-Wafa, Matthias Engelmann, Jessica Ezri, Christian Felley, Markus Fliegner, Nicolas Fournier, Montserrat Fraga, Yannick Franc, Pascal Frei, Remus Frei, Michael Fried, Florian Froehlich, Raoul Ivano Furlano, Luca Garzoni, Martin Geyer, Laurent Girard, Marc Girardin, Delphine Golay, Ignaz Good, Ulrike Graf Bigler, Beat Gysi, Johannes Haarer, Marcel Halama, Janine Haldemann, Pius Heer, Benjamin Heimgartner, Beat Helbling, Peter Hengstler, Denise Herzog, Cyrill Hess, Roxane Hessler, Klaas Heyland, Thomas Hinterleitner, Claudia Hirschi, Petr Hruz, Pascal Juillerat, Carolina Khalid-de Bakker, Stephan Kayser, Céline Keller, (Christina Knellwolf (-Grieger)), Christoph Knoblauch, Henrik Köhler, Rebekka Koller, Claudia Krieger(-Grübel), Patrizia Künzler, Rachel Kusche, Frank Serge Lehmann, Andrew Macpherson, Michel H. Maillard, Michael Manz, Astrid Marot, Rémy Meier, Christa Meyenberger, Pamela Meyer, Pierre Michetti, Benjamin Misselwitz, Patrick Mosler, Christian Mottet, Christoph Müller, Beat Müllhaupt, Leilla Musso, Michaela Neagu, Cristina Nichita, Jan Niess, Andreas Nydegger, Nicole Obialo, Diana Ollo, Cassandra Oropesa, Ulrich Peter, Daniel Peternac, Laetitia Marie Petit, Valérie Pittet, Daniel Pohl, Marc Porzner, Claudia Preissler, Nadia Raschle, Ronald Rentsch, Alexandre Restellini, Sophie Restellini, Jean-Pierre Richterich, Frederic Ris, Branislav Risti, Marc Alain Ritz, Gerhard Rogler, Nina Röhrich, Jean-Benoît Rossel, Vanessa Rueger, Monica Rusticeanu, Markus Sagmeister, Gaby Saner, Bernhard Sauter, Mikael Sawatzki, Michael Scharl, Martin Schelling, Susanne Schibli, Hugo Schlauri, Dominique Schluckebier, Daniela Schmid, Sybille Schmid (-Uebelhart), Jean-François Schnegg, Alain Schoepfer, Vivianne Seematter, Frank Seibold, Mariam Seirafi, Gian-Marco Semadeni, Arne Senning, Christiane Sokollik, Joachim Sommer, Johannes Spalinger, Holger Spangenberger, Philippe Stadler, Peter Staub, Dominic Staudenmann, Volker Stenz, Michael Steuerwald, Alex Straumann, Bruno Strebel, Andreas Stulz, Michael Sulz, Aurora Tatu, Michela Tempia-Caliera, Joël Thorens, Kaspar Truninger, Radu Tutuian, Patrick Urfer, Stephan Vavricka, Francesco Viani, Jürg Vögtlin, Roland Von Känel, Dominique Vouillamoz, Rachel Vulliamy, Paul Wiesel, Reiner Wiest, Stefanie Wöhrle, Samuel Zamora, Silvan Zander, Tina Wylie, Jonas Zeitz, Dorothee Zimmermann.

REFERENCES

1. Strober, W., Fuss, I.J. & Blumberg, R.S. The immunology of mucosal models of inflammation. *Annu Rev Immunol* **20**, 495-549 (2002).
2. Blumberg, R. & Powrie, F. Microbiota, disease, and back to health: a metastable journey. *Sci Transl Med* **4**, 137rv137 (2012).
3. Hegazy, A.N.*, et al.* Circulating and Tissue-Resident CD4(+) T Cells With Reactivity to Intestinal Microbiota Are Abundant in Healthy Individuals and Function Is Altered During Inflammation. *Gastroenterology* **153**, 1320-1337 e1316 (2017).
4. Maloy, K.J. & Powrie, F. Intestinal homeostasis and its breakdown in inflammatory bowel disease. *Nature* **474**, 298-306 (2011).
5. Rutgeerts, P.*, et al.* Effect of faecal stream diversion on recurrence of Crohn's disease in the neoterminal ileum. *Lancet* **338**, 771-774 (1991).
6. Zachos, M., Tondeur, M. & Griffiths, A.M. Enteral nutritional therapy for induction of remission in Crohn's disease. *Cochrane Database Syst Rev*, CD000542 (2007).
7. Cummings, J.H. & Kong, S.C. Probiotics, prebiotics and antibiotics in inflammatory bowel disease. *Novartis Found Symp* **263**, 99-111; discussion 111-114, 211-118 (2004).
8. Gecse, K.B.*, et al.* A global consensus on the classification, diagnosis and multidisciplinary treatment of perianal fistulising Crohn's disease. *Gut* **63**, 1381-1392 (2014).
9. Faith, J.J.*, et al.* The long-term stability of the human gut microbiota. *Science* **341**, 1237439 (2013).
10. Arumugam, M.*, et al.* Enterotypes of the human gut microbiome. *Nature* **473**, 174-180 (2011).
11. Schloissnig, S.*, et al.* Genomic variation landscape of the human gut microbiome. *Nature* **493**, 45-50 (2013).
12. Flint, H.J., Duncan, S.H. & Louis, P. The impact of nutrition on intestinal bacterial communities. *Curr Opin Microbiol* **38**, 59-65 (2017).
13. Falony, G.*, et al.* Population-level analysis of gut microbiome variation. *Science* **352**, 560-564 (2016).
14. Vandeputte, D.*, et al.* Stool consistency is strongly associated with gut microbiota richness and composition, enterotypes and bacterial growth rates. *Gut* **65**, 57-62 (2016).
15. Jostins, L.*, et al.* Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119-124 (2012).
16. Zuk, O., Hechter, E., Sunyaev, S.R. & Lander, E.S. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc Natl Acad Sci U S A* **109**, 1193-1198 (2012).
17. Uhlig, H.H. Monogenic diseases associated with intestinal inflammation: implications for the understanding of inflammatory bowel disease. *Gut* **62**, 1795-1805 (2013).
18. Lee, J.C. & Lennard-Jones, J.E. Inflammatory bowel disease in 67 families each with three or more affected first-degree relatives. *Gastroenterology* **111**, 587-596 (1996).
19. Lee, J.C.*, et al.* Gene expression profiling of CD8+ T cells predicts prognosis in patients with Crohn disease and ulcerative colitis. *J Clin Invest* **121**, 4170-4179 (2011).
20. Lee, J.C.*, et al.* Genome-wide association study identifies distinct genetic contributions to prognosis and susceptibility in Crohn's disease. *Nat Genet* **49**, 262-268 (2017).
21. Park, S.H., Aniwan, S. & Loftus, E.V., Jr. Advances in the use of biologics and other novel drugs for managing inflammatory bowel disease. *Curr Opin Pharmacol* **37**, 65-71 (2017).
22. Pillai, N., Dusheiko, M., Burnand, B. & Pittet, V. A systematic review of cost-effectiveness studies comparing conventional, biological and surgical interventions for inflammatory bowel disease. *PLoS One* **12**, e0185500 (2017).
23. Engels, M., Cross, R.K. & Long, M.D. Exercise in patients with inflammatory bowel diseases: current perspectives. *Clin Exp Gastroenterol* **11**, 1-11 (2018).
24. Eckburg, P.B.*, et al.* Diversity of the human intestinal microbial flora. *Science* **308**, 1635-1638 (2005).
25. Literature. Meta-analysis of IBD studies. *Various Journals* (2003 - 2017).
26. Png, C.W.*, et al.* Mucolytic bacteria with increased prevalence in IBD mucosa augment in vitro utilization of mucin by other bacteria. *Am J Gastroenterol* **105**, 2420-2428 (2010).
27. Sokol, H.*, et al.* Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 16731-16736 (2008).
28. Magnusdottir, S.*, et al.* Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nat Biotechnol* **35**, 81-89 (2017).
29. Fujimoto, T.*, et al.* Decreased abundance of Faecalibacterium prausnitzii in the gut microbiota of Crohn's disease. *J Gastroenterol Hepatol* **28**, 613-619 (2013).

30. Lopez-Siles, M.*, et al.* Mucosa-Associated Faecalibacterium prausnitzii Phylotype Richness Is Reduced in Patients with Inflammatory Bowel Disease. *Applied and environmental microbiology* **81**, 7582-7592 (2015).
31. Jostins, L.*, et al.* Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119-124 (2012).
32. Imhann, F.*, et al.* Interplay of host genetics and gut microbiota underlying the onset and clinical presentation of inflammatory bowel disease. *Gut* (2016).
33. Goodman, A.L. & Gordon, J.I. Our unindicted coconspirators: human metabolism from a microbial perspective. *Cell Metab* **12**, 111-116 (2010).
34. Muegge, B.D.*, et al.* Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science* **332**, 970-974 (2011).
35. Mainali, K.P.*, et al.* Statistical analysis of co-occurrence patterns in microbial presence-absence datasets. *PloS one* **12**, e0187132 (2017).
36. Williams, R.J., Howe, A. & Hofmockel, K.S. Demonstrating microbial co-occurrence pattern analyses within and between ecosystems. *Frontiers in Microbiology* **5**(2014).
37. Rahnavard, G.F., Eric, A.; McIver, Lauren, J.; Schwager, Emma; Weingart, George, Moon, Yo Sup ;Morgan, Xochitl C. ; Waldron, Levi; Huttenhower, Curtis High-sensitivity pattern discovery in large multi'omic datasets. (2017).
38. Plevinsky, J.M., Wojtowicz, A.A., Poulopoulos, N., Schneider, K.L. & Greenley, R.N. Perceived Impairment in Sports Participation in Adolescents With Inflammatory Bowel Diseases: A Preliminary Examination. *J Pediatr Gastroenterol Nutr* **66**, 79-83 (2018).
39. Torres, J., Mehandru, S., Colombel, J.F. & Peyrin-Biroulet, L. Crohn's disease. *Lancet* **389**, 1741-1755 (2017).
40. Khanna, R.*, et al.* Early combined immunosuppression for the management of Crohn's disease (REACT): a cluster randomised controlled trial. *Lancet* **386**, 1825-1834 (2015).
41. Ungaro, R., Mehandru, S., Allen, P.B., Peyrin-Biroulet, L. & Colombel, J.F. Ulcerative colitis. *Lancet* **389**, 1756-1770 (2017).
42. Magnusson, M.K.*, et al.* Anti-TNF Therapy Response in Patients with Ulcerative Colitis Is Associated with Colonic Antimicrobial Peptide Expression and Microbiota Composition. *J Crohns Colitis* **10**, 943-952 (2016).
43. Ghouri, Y.A.*, et al.* Systematic review of randomized controlled trials of probiotics, prebiotics, and synbiotics in inflammatory bowel disease. *Clin Exp Gastroenterol* **7**, 473-487 (2014).
44. Topping, D.L. & Clifton, P.M. Short-chain fatty acids and human colonic function: roles of resistant starch and nonstarch polysaccharides. *Physiol Rev* **81**, 1031-1064 (2001).
45. Arpaia, N.*, et al.* Metabolites produced by commensal bacteria promote peripheral regulatory T-cell generation. *Nature* **504**, 451-455 (2013).
46. Wu, F.*, et al.* Phascolarctobacterium faecium abundant colonization in human gastrointestinal tract. *Exp Ther Med* **14**, 3122-3126 (2017).
47. Chen, J.*, et al.* Dysbiosis of intestinal microbiota and decrease in paneth cell antimicrobial peptide level during acute necrotizing pancreatitis in rats. *PLoS One* **12**, e0176583 (2017).
48. Hancock, L. & Mortensen, N.J. How Often Do IBD Patients Require Resection of Their Intestine? *Inflammatory Bowel Diseases* **14**, S68-S69 (2008).
49. Olivera, P., Spinelli, A., Gower-Rousseau, C., Danese, S. & Peyrin-Biroulet, L. Surgical rates in the era of biological therapy: up, down or unchanged? *Curr Opin Gastroenterol* **33**, 246-253 (2017).
50. Rutgeerts, P.*, et al.* Predictability of the postoperative course of Crohn's disease. *Gastroenterology* **99**, 956-963 (1990).
51. Patterson, A.M.*, et al.* Human Gut Symbiont Roseburia hominis Promotes and Regulates Innate Immunity. *Front Immunol* **8**, 1166 (2017).
52. Machiels, K.*, et al.* Specific members of the predominant gut microbiota predict pouchitis following colectomy and IPAA in UC. *Gut* **66**, 79-88 (2017).
53. Machiels, K.*, et al.* A decrease of the butyrate-producing species Roseburia hominis and Faecalibacterium prausnitzii defines dysbiosis in patients with ulcerative colitis. *Gut* **63**, 1275-1283 (2014).
54. Flint, H.J., Duncan, S.H., Scott, K.P. & Louis, P. Links between diet, gut microbiota composition and gut metabolism. *Proc. Nutr. Soc.* **74**, 13-22 (2015).
55. Byndloss, M.X.*, et al.* Microbiota-activated PPAR-gamma signaling inhibits dysbiotic Enterobacteriaceae expansion. *Science* **357**, 570-575 (2017).

MAIN FIGURE LEGENDS

Figure 1. Distinct microbial features associated with IBD patients. (a,b) Microbial clustering is shown based on Bray-Curtis dissimilarity principal coordinate analysis (PCoA) metrics using 2-4 intestinal biopsy samples per patient. 346 patients with IBD in *Cohort 1* **(a)** and 156 patients with IBD and 227 non-IBD subjects in *Cohort 2* **(b)** were included into the study. Non-parametric analysis of variance (Adonis) was used to test significant difference between groups on PCoA plot; in each case $p<0.01$ for each disease group in each cohort and non-parametric two-sided Mann-Whitney U-test was used to analyze significant differences between ranks ± s.d for PC1 and PC2. $p<0.05$ was considered significant. Ellipsoids represent a 95% confidence interval surrounding each disease group. Box-and-whisker plots display 1st and 3rd quartiles and whiskers go from each quartile to the minimum or maximum. Significant comparisons are shown on the figure (b). Groups are not significantly different among each other on PC2 axis (b). **(c)** Overall representation of overlapping significant taxa of disease groups between *Cohort 1* and *Cohort 2*. **(d, e)** Taxa significantly associated with "active" or "quiescent" clinical course in each cohort are plotted as relative abundance ratios in CD (**d)** and UC **(e)** disease groups. 494 CD and 447 UC samples in *Cohort 1* (d) and 226 CD and 195 UC in *Cohort 2* (e) were analyzed. Statistical analysis for taxonomy results were performed using MaAsLin pipeline with BH-FDR (q value) (c-e). $q<0.05$ was considered significant and p value showing the trend ($p<0.05$ and $q<0.2$) and are indicated in the figures (d,e). Taxa with open bars show opposite relative abundance ratio between cohorts. • denotes classified but unnamed genus, ••denotes unclassified genus, and ••• denotes taxa are recommended under different family based on genome trees in Greengenes database.

Figure 2. Microbial co-occurrence analysis of IBD and non-IBD subjects with GI ecosystems. (a,b) Ecosystem-specific co-occurrence patterns at genus level are visualized using network diagrams where microbial taxa represent nodes; the presence of a positive co-occurrence relationship based on correlation is represented by an edge for *Cohort 1* **(a)** and *Cohort 2* **(b)**. Co-occurrence relationships with strong Spearman's correlation coefficient (ρ-value > 0.5 and $p<0.05$) are depicted with network diagram for each disease. Each cluster was assigned with a subscript to related disease or non-IBD control *(e.g.* CD_A, UC_A, $Cont_A$).

Figure 3. Clinical determinants correlate with intestinal microbiota composition of IBD patients. (a-d) Hierarchical analysis performed using end-to-end statistical method (HAllA) at phylum **(a, c)** and genus rank **(b, d)** for *Cohort 1* (**a,b)** and *Cohort 2* **(c,d)**. Significant metadata among many clinical parameters was plotted after BH-FDR correction. $q<0.05$ was considered significant (a-d). Association rank was sorted by high similarity score. The color intensity of the heat-map (pairwise normalized mutual information [NMI]) and numbers in cells identify significant pairs of features (clinical determinants vs. microbial composition) of IBD patients.

Figure 4. Intestinal microbiota composition of IBD patients with responsiveness to anti-TNF-α or corticosteroid therapies. (a) Microbial clustering is shown based on Bray-Curtis dissimilarity PCoA metrics of intestinal biopsy samples from IBD patients responding or non-responding to anti-TNF-α therapy showing reproducibility in both cohorts. CD (*solid line*) and UC (*dashed line*) are used to identify the disease groups in PCoA plot. Ellipsoids represent a 95% confidence interval surrounding each group. Adonis test assessed the significant difference between groups in PCoA plot: $p < 0.01$ is for each compared group in both cohorts. **(b)** Overall representation of overlapping significant taxa of responding and non-responding groups in each cohort are shown using phylogenetic trees for anti-TNF-α therapy. Closed circles correspond to *Cohort 1* and open circles correspond to *Cohort 2*, green coloration represents higher relative

abundances in responding groups and red coloration represents higher relative abundance in non-responding groups. Significant taxa are plotted with respective circles. **(c)** Significant unique microbial taxa identified as a signature of responding (success) and non-responding (failure) groups CD and UC are shown for *Cohort 1* and *Cohort 2* after using MaAsLin with BH-FDR (q value). **(d)** Overall representation of overlapping significant taxa of responding and non-responding groups in each cohort are shown using phylogenetic trees for corticosteroid therapy. q<0.05 was considered significant and are plotted (b-d). 88 success and 91 failure CD samples with 19 success and 48 failure UC samples in *Cohort 1* and 129 success and 27 failure CD samples with 53 success and 18 failure UC samples in *Cohort 2* were analyzed (a-d).

Figure 5. Disease locations are critically important determinants that correlate with intestinal microbiota of IBD patients. (a,b) Microbial composition differences in IBD patients within different disease locations in *Cohort 1* **(a)** and *Cohort 2* **(b)** are plotted using beta diversity (Bray-Curtis dissimilarity PCoA) and alpha diversity (Shannon and Simpson index). Ellipsoids in PCoA plots represent a 95% confidence interval surrounding each group and Box-and-whisker plots display 1st and 3rd quartiles and whiskers go from each quartile to the minimum or maximum. There is significant separation between groups on PCoA plot (p<0.05). Significant differences between groups in alpha diversity analysis were determined by an ordinary one-way ANOVA corrected for multiple comparisons using BH-FDR. **(c)** Only significant bacterial abundance differences between colonic CD, ileal CD and colonic UC samples are plotted using the samples identified in (a,b). Differences between tested groups were determined using MaAsLin with BH-FDR correction (q value). Asterisk indicates the similar significant taxonomic changes observed in *Cohort 2*. **(d,e)** Only significantly different metabolic subsystems in colonic CD samples **(d)** and ileal involved CD samples **(e)** when compared with UC samples are shown. q<0.05 was considered significant (c-e) and error bars on box plots are standard deviations (d,e). The same samples described in (a,b) were used for analysis of metabolic subsystems (d,e). 370 ileal, 110 colonic and 447 UC in *Cohort 1* and 203 ileal, 28 colonic and 221 UC in *Cohort 2* were used for analysis (a-e).

Figure 6. Altered microbiota profiles in inactive CD patients following surgical resection of active disease segments. (a-d) Species richness between CD patients following surgery or without surgical resection was calculated using alpha diversity (Shannon and Simpson indices) for *Cohort 1* **(a)** and *Cohort 2* **(b)**. Box-and-whisker plots display 1st and 3rd quartiles and whiskers go from each quartile to the minimum or maximum. Microbial composition differences between these samples was calculated with PCoA on Bray-Curtis distance for *Cohort 1* **(c)** and *Cohort 2* **(d)**. Blue color corresponds to CD patients having undergone surgical resection and red color corresponds to CD patients without surgery. Ellipsoids represent a 95% confidence interval surrounding each group (c,d). Non-parametric two-sided Mann-Whitney U-test and Adonis test were used to identify the statistical significant differences between groups for alpha (a,b) and beta diversity (c,d), respectively. There is a significant separation on PCoA plots between compared groups analyzed in (c) and (d) (p<0.05) **(e, f)** Only significant taxa (q<0.05) associated with surgical resection status are plotted as relative abundance ratio for CD in *Cohort 1* **(e)** and *Cohort 2* **(f)**. In *Cohort 2*, non-significant data (q>0.05) was shown with "ns" (f). 105 (+) and 109 (-) samples in *Cohort 1* (a) and 96 (+) and 371 samples (-) in *Cohort 2* were analyzed (a-f).

METHODS

Cohorts and study design

Biopsy samples in this study were collected either a) in the Swiss IBD cohort study[124] which has included more than 3500 patients (*Cohort 1*; approximately 5000 intestinal biopsies have been submitted to the Biobank Cohort) or b) newly recruited IBD patients and non-IBD subjects in the framework of the human intestinal community project of Bern University Hospital (*Cohort 2*). *Cohort 1* includes biopsy samples from teaching hospitals, community hospitals and office practices in Switzerland. For all patients included in the cohort, deep longitudinal clinical and investigational phenotyping for an average of 5.7 years (range 1-11 years) was sequentially taken over at least annual follow-ups. Unlike the Swiss IBD cohort, which had focused exclusively on inflammatory bowel disease, *Cohort 2* included asymptomatic control patients undergoing screening colonoscopies without macroscopic or microscopic abnormalities, in whom all measured hematologic and clinical chemistry parameters were normal. Non-IBD biopsy samples were collected from subjects that were registered for a screening ileo-colonoscopy without any gastrointestinal symptoms, no suspected functional intestinal symptoms and where all other biochemical and hematological work-up was negative.

Inclusion criteria for Swiss IBD cohort patients were on established Lennard-Jones criteria[125], confirmed by radiological, endoscopic and histological findings. Subject to informed consent for the longitudinal study, non-residents of Switzerland and patients unable or unwilling to provide blood samples were excluded. At inclusion, the patient's physician filled an enrolment questionnaire to make a clinical evaluation of the disease activity (CDAI, HBI for CD and MTWAI, SCCAI for UC), the disease extent and location based on the Montréal classification, including extraintestinal manifestations, previous surgery (bowel resection or perianal), complications (fistula, stenosis, anaemia, osteoporosis), family history, duration of disease, smoking status, previous treatment success/failure and current therapy and dosage, as well as laboratory data (C-reactive protein, hemoglobin, auto-antibodies, from 1244 patients also stool calprotectin). Clinical treatment for IBD was scored (5-ASA preparations, steroids, thiopurines, methotrexate, tumor necrosis factor α (TNF-α) inhibitors or other biologicals) as well as the use of other concurrent medications and non-IBD surgeries. The clinical follow-up was on an annual basis and prospectively collected by the clinician or a study nurse with an additional questionnaire about their psychosocial situation (quality of life, depression and anxiety) and clinical IBD trajectory (disease activity, hospitalization, change in medication, adherence, surgeries). Hard copy data from the enrolment and clinical follow-up were then digitalized using Access 2010 (Microsoft, USA).

The inclusion criteria and data collection for consenting patients for the local Bern cohort were similar, except that patients were diagnosed and followed up in the Bern University Hospital (Inselspital) IBD clinic and in the Bern City Hospitals.

Clinical data extraction, categorization and analysis

Structured clinical metadata was prospectively collected according to pre-determined standards containing variables of *Cohort 1* and *Cohort 2* (documented electronically using Research Electronic Data Capture (REDCap) database[126]). These metadata were imported and processed as *dataframe* in R

(http://www.r-project.org) using *xlsx* package. Assessment of the microbiota composition from intestinal biopsies was then analyzed according to parameters including the type, course and severity of disease using the detailed clinical criteria summarized in Supplementary Table 1 for *Cohort 1* and Supplementary Table 2 for *Cohort 2*. Statistical analyses were performed using Student's t-test, Wilcoxon's rank sum test and Pearson's chi-squared test to test normality between groups.

Ethics statement

Licensed gastroenterologists collected biopsy samples and clinical data of patients identified as CD or UC and non-IBD subjects. The SIBDC study (Cohort 1) protocol was approved by the Ethics Committee of the Kanton Zurich (reference KEK-ZH-number 2013-0284) and this cohort is also registered with Trials Registration Number NCT02849821. The Bern Human Intestinal Community project (Cohort 2) was approved by the Bern Cantonal Ethics Commission (Ref: KEK-BE: 251/14 and 336/14) with signed informed consent was obtained from all participants. Bern cohort data were anonymized and collected in EDC (electronic data capture) system (REDCap)[126], activated for the trial after successfully passing formal quality controls. The Bern EDC system and the database are hosted by the Clinical Trial Unit of Bern University. The study complies with all relevant ethical regulations.

DNA extraction from human biopsies

Intestinal endoscopic biopsies were initially collected into 2ml microfuge tubes containing 500-600 µl RNAlater (Sigma-Aldrich) and stored at -20°C prior to DNA extraction protocol[127]. Total DNA was isolated using AllPrep DNA/RNA Mini Kit (Qiagen) according to the manufacturer's instructions. Briefly, 600 µl of Buffer RLT Plus 2-mercaptoethanol and a metal bead were added into each tube. Samples were then homogenized with the Retsch Tissue Lyser (QIAGEN) at 30Hz for 3 min, followed by 3 min centrifugation at 13 000g (Eppendorf). Supernatants were transferred into the AllPrep DNA mini spin column and centrifuged at 9 000g for 30 sec. DNA attached to spin columns was washed/de-salted using 500 µl of Buffer AW1 and Buffer AW2. Lastly, DNA samples were eluted with 25-30 µl EB buffer into 1.5 ml microfuge tubes. The concentration and purity of the isolated DNA samples were analyzed in NanoDrop® (Thermo Scientific).

Microbial profiling of biopsies

1255 samples from *Cohort 1* and 1846 biopsy samples from *Cohort 2* were used in this study. The V5/V6 region of 16S rRNA genes was amplified with Invitrogen™ Platinum™ *Taq* DNA polymerase from 1000-2000 ng of DNA using a range of oligonucleotide primers specific for the domains V5 and V6 of rDNA bacteria. Specifically, all forward core primers had been modified by the addition of a PGM sequencing adaptor, a 'GT' spacer and unique barcode that allowed us to have up to different 96 barcodes. The expected product length was ~350 bp including adaptors and barcodes. Bacteria-specific primers (forward 5'-CCATCTCATCCCTGCGTGTCTCCGACTCAGC-barcode-ATTAGATACCCYGGTAGTCC-3' and reverse 5'-CCTCTCTATGGGCAGTCGGTGATACGAGCTGACGACARCCATG-3') were used[128,129]. PCR conditions consisted of an initial 5 min at 94°C denaturation step, followed by 35 cycles of 1 min denaturation at 94°C,

20s annealing cycle at 46°C and 30s extension cycle at 72°C, with a final extension for 7 min at 72°C. Samples were kept at 4°C until loading into an 1% agarose gel. Amplicons were purified using Gel Extraction Kit (Qiagen) after gel electrophoresis for 1 hour. The concentrations of amplicons were then evaluated by Qubit 3.0 Fluorometer (ThermoFisher) prior to library preparation and used at 26pM. To prepare template-positive Ion PGM™ Template OT2 400 Ion Sphere™ Particles (ISPs) containing clonally amplified DNA we used the Ion OneTouch™ Instrument with the Ion PGM™ Template OT2 400 Kit (ThermoFisher). Sequencing was performed using the Ion PGM™ Sequencing 400 Kit and Ion 316™ Chip V2 within the Ion PGM™ System (Thermo Fisher)[130].

Raw sequences were first loaded into the QIIME 1.9.1 pipeline, as described[131] using custom analysis scripts for analysis on the UBELIX Linux cluster of the University of Bern. Only samples with more than 4500 high quality reads were then used for further analysis using R packages. Operational taxonomic units were picked using UCLUST with a 97% sequence identity threshold, using the default options as implemented in QIIME and followed by taxonomy assignment using the latest Greengenes database (greengenes.secondgenome.com). OTUs that were not present in at least 0.01% of our samples or with a low abundance (<0.001% of the total counts) were filtered out. In combination with the detailed clinical patient metadata (diagnosis, age, BMI, inflammation status, location of biopsy, disease severity, treatment etc.) Principal Coordinates Analysis were performed to determine whether samples with certain properties cluster. Due to the limitations of the resolution on taxonomical classification using 16S gene sequencing technique, analyses were restricted to genus level. The OTU abundance *biom* file and mapping file were used for statistical analyses and data visualization in the R statistical programming environment package *phyloseq*[132]. Calculation of the α-diversity (Observed OTUs, Simpson and Shannon index), β-diversity (Bray-Curtis genus-level community dissimilarities, weighted and unweighted UniFrac-based PCoA), and statistical analysis of clustering using Mann-Whitney U tests for alpha diversity and Adonis (PERMANOVA) for beta diversity to confirm that the strength and statistical significance of groups in the same distance metrics in the QIIME pipeline and *phyloseq* in R[132,133]. Multivariate homogeneity of group dispersion was to calculate the average distance of the groups and to test if the dispersion of any group was significantly different from the others. Using k-nearest neighborhood networks created by thresholding the Jaccard dissimilarity matrix, the differences in microbial community diversities between biopsies collected from different locations (ileum, colon or rectum) for both cohorts were analyzed. Further analyses were performed using a Random Forest supervised learning algorithm with additional Boruta feature selection to identify the most important predictive taxa or community within microbial communities in R. Hierarchical All-against-All significance testing (HAllA)[134] and multivariate analysis by linear models (MaAsLin) R package[135] were used to find associations between clinical metadata (age, BMI, gender, smoking habits, medication, inflammation status, anatomic location etc.) and microbial community abundance. The Q-value package was implemented in MaAsLin to correct for multiple testing (Benjamini-Hochberg false discovery rate correction; a false discovery rate [FDR], q-value) of 0.05. Taxa present in at least 30% of the samples and taxa that had more than 0.0001% of total abundance were set as the cut-off values for further analysis. After correction for a false discovery rate, $q<0.05$ was considered significant. Notable near-significant differences ($0.05<p$ and $0.05<q<0.2$) were also highlighted in the figures. Non-significant data were indicated in figures by “n.s”.

All the relevant codes for running the MaAsLin and HAllA pipeline are available in Dr. Huttenhower's group webpage (https://huttenhower.sph.harvard.edu). Plots were generated with ggplot2 using *phyloseq* object or Graphpad Prism v7.0a.

Biopsies (1255 samples in *Cohort 1* and 1846 in *Cohort 2*) from different pre-defined positions along the length of the colon and the distal small intestine (terminal ileum) from each patient generated 35,678,215 and 40,879,931 high-quality 16S rDNA reads with an average amplicon length of 293bp (Supplementary Tables 1 and 2). Sequences were clustered into 359,010 and 392,170 taxa respectively based on their shared sequence similarity at a 97% threshold (3% sequence divergence)[136]. The average estimated coverage for OTU characterization was 94.2% and 93.2% [136,137].

Microbial association network construction within disease status

Co-occurrence analyses were performed based on the organization of the data generated by *phyloseq* object in R. Network analyses were performed using the scripts reported by Williams et al.[138]. Briefly, we considered the co-occurrence to be positive rank correlations identified using Spearman's correlation between identified taxa within each disease condition. Positive correlation (weak correlation: ρ-value > 0.25 and strong correlation: ρ-value > 0.5) helped us to identified even relatively less strong and highly correlated significant positive correlations between co-occurring taxa in each identified ecosystem, respectively. The OTUs that had more than 0.0001% total of relative abundance and taxa present in at least 30% of the samples were retained for further statistical analysis and illustration of the significant co-occurrence network between microbial taxa. We thus obtained a data frame that listed disease status, the originating environment of the samples, assigned taxa, Spearman's correlation coefficient, p value calculated using PERMANOVA, and the abundances of each taxa. We calculated the modules detected using Vertex and edge betweenness centrality functions in *igraph* package and generated betweenness score for each microbial taxon. The 'betweenness centrality scores' for each vertex is the number of these shortest paths going through a node and 'eigenvector centrality' is a measure of node connectivity to important neighbors in a given network. Spearman's correlation coefficients calculated between the different co-occurring taxa within each environment were used for generating network plots. *Igraph* objects were transferred into Cytoscape using *RJSONIO* and *httr* packages in R. The community structure analysis of biological co-occurrence network carried out using GLay implemented into Cytoscape as a fast-greedy algorithm[139]. Network plots were generated using a customized organic layout of Cytoscape *v.3.5.1*.

Reaction-level analysis

OTUs were mapped to metabolic reactions via a previously published collection of genome-scale metabolic models (GSMMs)[140]. We included OTUs that could be mapped to a taxonomic rank of family or lower and to at least one GSMM. The normalized abundance of a reaction i in a patient sample, $a_r(i)$ was calculated as:

$$a_r(i) = \sum_{j=1}^{N} a_{otu}(j)E(i,j) \Big/ \sum_{j=1}^{N} a_{otu}(j),$$

where $a_{otu}(j)$ is the abundance of OTU j in the patient sample, N is the number of detected OTUs, and $E(i,j)$ is the expected probability (frequency of occurrences) of reaction i in the GSMMs associated with OTU j. To analyze metabolic reaction differences between patient groups, a two-sample *t*-test was used to compare the mean abundances of each reaction and each binary patient classification of interest (e.g., UC vs. CD) between the two groups. The Benjamini-Hochberg procedure[141] was applied to correct for multiple testing and adj-*p*-value ≤ 0.05 between groups was considered significant. The metabolic subsystem classification of reactions was obtained from the GSMMs and for each *t*-test comparing groups of patients and each subsystem, Fisher's exact test was performed to determine if the subsystem was overrepresented among the significantly different reactions; subsystems with p-value ≤ 0.05 were considered enriched. All analyses were performed in MATLAB (R2018a).

Fecal calprotectin

Calprotectin is a member of the S100 family of calcium binding proteins found in the cytosol of neutrophils and thus, abundant in all body fluids in proportion to the degree of inflammation present. Fecal calprotectin was measured by a quantitative enzyme-linked immunosorbent assay after using Calprotectin – Stool Extraction with CALEX. The assays were carried out according to the instructions (The BÜHLMANN fecal Calprotectin Assays, BÜHLMANN Laboratories AG, Schönenbuch, Switzerland). All fecal samples were processed within 72h after collection. The assay plates were read at an OD_{450} nm as instructed in manufacture's documentation. The calprotectin cut-off level representing a positive value was ≥ 50µg calprotectin/gr stool.

Search Strategy and Data Extraction

We systematically searched the publications in PubMed (including MEDLINE), and Google Scholar. The electronic search algorithm consisted of terms relating to key concepts of "inflammatory bowel disease", "IBD", "microbiome", "gut microbiota", "metagenomic", "16S rDNA", "human", "mice", "cats", "dogs", "animal models", "naturally induced colitis" and "chemically induced colitis". Reference lists of included articles, related reviews, and other relevant sources were manually searched. After initial screening of titles and abstracts, full text assessments were performed for eligibility of the data to include into the figure. Characteristics of each eligible study were extracted: disease information, species, race (ethnicity), gender, age (median year), number of samples used for sequencing, type of samples, method of characterization of the microbiota, model of inflammation induction. Results of included studies were then organized, and corresponding heat-maps generated using Euclidean clustering of sample annotations (vertical) including race/ethnicity, gender, median age, patient number, sample type, model of inflammation induction, sequencing method and microbial taxa (horizontal) at different taxonomic ranks.

Reporting Summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Method References

1. Liguori, G., et al. Fungal Dysbiosis in Mucosa-associated Microbiota of Crohn's Disease Patients. J Crohns Colitis 10, 296-305 (2016).
2. Papa, E., et al. Non-invasive mapping of the gastrointestinal microbiota identifies children with inflammatory bowel disease. PloS one 7, e39242 (2012).
3. Santoru, M.L., et al. Cross-sectional evaluation of the gut-microbiome metabolome axis in an Italian cohort of IBD patients. Sci Rep 7, 9523 (2017).
4. Pedamallu, C.S., et al. Metagenomic Characterization of Microbial Communities In Situ Within the Deeper Layers of the Ileum in Crohn's Disease. Cell Mol Gastroenterol Hepatol 2, 563-566 e565 (2016).
5. Kellermayer, R., et al. Microbiota separation and C-reactive protein elevation in treatment-naive pediatric granulomatous Crohn disease. J Pediatr Gastroenterol Nutr 55, 243-250 (2012).
6. Frank, D.N., et al. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. Proceedings of the National Academy of Sciences of the United States of America 104, 13780-13785 (2007).
7. Willing, B.P., et al. A pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes. Gastroenterology 139, 1844-1854 e1841 (2010).
8. Wang, W., et al. Increased proportions of Bifidobacterium and the Lactobacillus group and loss of butyrate-producing bacteria in inflammatory bowel disease. Journal of clinical microbiology 52, 398-406 (2014).
9. Rajilic-Stojanovic, M., Shanahan, F., Guarner, F. & de Vos, W.M. Phylogenetic analysis of dysbiosis in ulcerative colitis during remission. Inflamm Bowel Dis 19, 481-488 (2013).
10. Seksik, P., et al. Alterations of the dominant faecal bacterial groups in patients with Crohn's disease of the colon. Gut 52, 237-242 (2003).
11. Halfvarson, J., et al. Dynamics of the human gut microbiome in inflammatory bowel disease. Nat Microbiol 2, 17004 (2017).
12. Erickson, A.R., et al. Integrated metagenomics/metaproteomics reveals human host-microbiota signatures of Crohn's disease. PloS one 7, e49138 (2012).
13. Eun, C.S., et al. Does the intestinal microbial community of Korean Crohn's disease patients differ from that of western patients? BMC Gastroenterol 16, 28 (2016).
14. Forbes, J.D., Van Domselaar, G. & Bernstein, C.N. Microbiome Survey of the Inflamed and Noninflamed Gut at Different Compartments Within the Gastrointestinal Tract of Inflammatory Bowel Disease Patients. Inflammatory Bowel Diseases 22, 817-825 (2016).
15. Andoh, A., et al. Multicenter analysis of fecal microbiota profiles in Japanese patients with Crohn's disease. J Gastroenterol 47, 1298-1307 (2012).
16. Rehman, A., et al. Transcriptional activity of the dominant gut mucosal microbiota in chronic inflammatory bowel disease patients. J Med Microbiol 59, 1114-1122 (2010).
17. Martinez-Medina, M., Aldeguer, X., Gonzalez-Huix, F., Acero, D. & Garcia-Gil, L.J. Abnormal microbiota composition in the ileocolonic mucosa of Crohn's disease patients as revealed by polymerase chain reaction-denaturing gradient gel electrophoresis. Inflamm Bowel Dis 12, 1136-1145 (2006).
18. Norman, J.M., et al. Disease-specific alterations in the enteric virome in inflammatory bowel disease. Cell 160, 447-460 (2015).
19. Ashton, J.J., et al. 16S sequencing and functional analysis of the fecal microbiome during treatment of newly diagnosed pediatric inflammatory bowel disease. Medicine 96(2017).
20. Kabeerdoss, J., Jayakanthan, P., Pugazhendhi, S. & Ramakrishna, B.S. Alterations of mucosal microbiota in the colon of patients with inflammatory bowel disease revealed by real time polymerase chain reaction amplification of 16S ribosomal ribonucleic acid. Indian J Med Res 142, 23-32 (2015).
21. Kolho, K.L., et al. Fecal Microbiota in Pediatric Inflammatory Bowel Disease and Its Relation to Inflammation. Am J Gastroenterol 110, 921-930 (2015).

22. Machiels, K., et al. A decrease of the butyrate-producing species Roseburia hominis and Faecalibacterium prausnitzii defines dysbiosis in patients with ulcerative colitis. Gut 63, 1275-1283 (2014).
23. Martinez-Medina, M., et al. Molecular diversity of Escherichia coli in the human gut: new ecological evidence supporting the role of adherent-invasive E. coli (AIEC) in Crohn's disease. Inflamm Bowel Dis 15, 872-882 (2009).
24. Duboc, H., et al. Connecting dysbiosis, bile-acid dysmetabolism and gut inflammation in inflammatory bowel diseases. Gut 62, 531-539 (2013).
25. Schwiertz, A., et al. Microbiota in pediatric inflammatory bowel disease. J Pediatr 157, 240-244 e241 (2010).
26. Kaakoush, N.O., et al. Microbial dysbiosis in pediatric patients with Crohn's disease. Journal of clinical microbiology 50, 3258-3266 (2012).
27. Walker, A.W., et al. High-throughput clone library analysis of the mucosa-associated microbiota reveals dysbiosis and differences between inflamed and non-inflamed regions of the intestine in inflammatory bowel disease. Bmc Microbiol 11(2011).
28. Gophna, U., Sommerfeld, K., Gophna, S., Doolittle, W.F. & van Zanten, S.J.O.V. Inerences between tissue-associated intestinal microfloras of patients with Crohn's disease and ulcerative colitis. Journal of clinical microbiology 44, 4136-4141 (2006).
29. Knoll, R.L., et al. Gut microbiota differs between children with Inflammatory Bowel Disease and healthy siblings in taxonomic and functional composition: a metagenomic analysis. Am J Physiol-Gastr L 312, G327-G339 (2017).
30. Hansen, R., et al. Microbiota of De-Novo Pediatric IBD: Increased Faecalibacterium Prausnitzii and Reduced Bacterial Diversity in Crohn's But Not in Ulcerative Colitis. American Journal of Gastroenterology 107, 1913-1922 (2012).
31. Jacobs, J.P., et al. A Disease-Associated Microbial and Metabolomics State in Relatives of Pediatric Inflammatory Bowel Disease Patients. Cell Mol Gastroenterol Hepatol 2, 750-766 (2016).
32. Bajer, L., et al. Distinct gut microbiota profiles in patients with primary sclerosing cholangitis and ulcerative colitis. World journal of gastroenterology : WJG 23, 4548-4558 (2017).
33. Baumgart, M., et al. Culture independent analysis of ileal mucosa reveals a selective increase in invasive Escherichia coli of novel phylogeny relative to depletion of Clostridiales in Crohn's disease involving the ileum. Isme J 1, 403-418 (2007).
34. Gevers, D., et al. The treatment-naive microbiome in new-onset Crohn's disease. Cell host & microbe 15, 382-392 (2014).
35. Tong, M., et al. A modular organization of the human intestinal mucosal microbiota and its association with inflammatory bowel disease. PloS one 8, e80702 (2013).
36. Pascal, V., et al. A microbial signature for Crohn's disease. Gut 66, 813-822 (2017).
37. Mottawea, W., et al. Altered intestinal microbiota-host mitochondria crosstalk in new onset Crohn's disease. Nat Commun 7, 13419 (2016).
38. Morgan, X.C., et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. Genome Biol 13, R79 (2012).
39. Verma, R., Verma, A.K., Ahuja, V. & Paul, J. Real-time analysis of mucosal flora in patients with inflammatory bowel disease in India. Journal of clinical microbiology 48, 4279-4282 (2010).
40. Arun Gupta, S.K., Josef Wagner, Carl Kirkwood, Mark Morrison, Chris McSweeney, and Finlay Macrae. Analysis of Mucosal Microbiota in In ammatory Bowel Disease using a Custom Phylogenetic Microarray. Austin Journal of Gastroenterology 1, 1-6 (2014).
41. Bibiloni, R., Mangold, M., Madsen, K.L., Fedorak, R.N. & Tannock, G.W. The bacteriology of biopsies differs between newly diagnosed, untreated, Crohn's disease and ulcerative colitis patients. Journal of Medical Microbiology 55, 1141-1149 (2006).
42. Sokol, H., et al. Low counts of Faecalibacterium prausnitzii in colitis microbiota. Inflamm Bowel Dis 15, 1183-1189 (2009).
43. Lopez-Siles, M., et al. Mucosa-Associated Faecalibacterium prausnitzii Phylotype Richness Is Reduced in Patients with Inflammatory Bowel Disease. Applied and environmental microbiology 81, 7582-7592 (2015).

44. Andoh, A., et al. Comparison of the fecal microbiota profiles between ulcerative colitis and Crohn's disease using terminal restriction fragment length polymorphism analysis. Journal of Gastroenterology 46, 479-486 (2011).
45. Swidsinski, A., Loening-Baucke, V., Vaneechoutte, M. & Doerffel, Y. Active Crohn's disease and ulcerative colitis can be specifically diagnosed and monitored based on the biostructure of the fecal flora. Inflammatory Bowel Diseases 14, 147-161 (2008).
46. Nishino, K., et al. Analysis of endoscopic brush samples identified mucosa-associated dysbiosis in inflammatory bowel disease. J Gastroenterol (2017).
47. Chen, L., et al. Characteristics of fecal and mucosa-associated microbiota in Chinese patients with inflammatory bowel disease. Medicine (Baltimore) 93, e51 (2014).
48. Suchodolski, J.S., Dowd, S.E., Wilke, V., Steiner, J.M. & Jergens, A.E. 16S rRNA gene pyrosequencing reveals bacterial dysbiosis in the duodenum of dogs with idiopathic inflammatory bowel disease. PloS one 7, e39333 (2012).
49. Xenoulis, P.G., et al. Molecular-phylogenetic characterization of microbial communities imbalances in the small intestine of dogs with inflammatory bowel disease. FEMS Microbiol Ecol 66, 579-589 (2008).
50. Omori, M., et al. Fecal microbiome in dogs with inflammatory bowel disease and intestinal lymphoma. J Vet Med Sci 79, 1840-1847 (2017).
51. Allenspach, K., et al. Evaluation of mucosal bacteria and histopathology, clinical disease activity and expression of Toll-like receptors in German shepherd dogs with chronic enteropathies. Veterinary microbiology 146, 326-335 (2010).
52. Suchodolski, J.S., Xenoulis, P.G., Paddock, C.G., Steiner, J.M. & Jergens, A.E. Molecular analysis of the bacterial microbiota in duodenal biopsies from dogs with idiopathic inflammatory bowel disease. Veterinary microbiology 142, 394-400 (2010).
53. Suchodolski, J.S., et al. The fecal microbiome in dogs with acute diarrhea and idiopathic inflammatory bowel disease. PloS one 7, e51907 (2012).
54. Rossi, G., et al. Comparison of microbiological, histological, and immunomodulatory parameters in response to treatment with either combination therapy with prednisone and metronidazole or probiotic VSL#3 strains in dogs with idiopathic inflammatory bowel disease. PloS one 9, e94699 (2014)
55. Minamoto, Y., et al. Alteration of the fecal microbiota and serum metabolite profiles in dogs with idiopathic inflammatory bowel disease. Gut microbes 6, 33-47 (2015).
56. Jergens, A.E. Feline idiopathic inflammatory bowel disease: what we know and what remains to be unraveled. J Feline Med Surg 14, 445-458 (2012).
57. Vazquez-Baeza, Y., Hyde, E.R., Suchodolski, J.S. & Knight, R. Dog and human inflammatory bowel disease rely on overlapping yet distinct dysbiosis networks. Nat Microbiol 1, 16177 (2016).
58. Inness, V.L., McCartney, A.L., Khoo, C., Gross, K.L. & Gibson, G.R. Molecular characterisation of the gut microflora of healthy and inflammatory bowel disease cats using fluorescence in situ hybridisation with special reference to Desulfovibrio spp. J Anim Physiol an N 91, 48-53 (2007).
59. Elinav, E., et al. NLRP6 inflammasome regulates colonic microbial ecology and risk for colitis. Cell 145, 745-757 (2011).
60. Li, M., et al. Upregulation of intestinal Barrier Function in Mice with Dss-induced colitis by a Defined Bacterial consortium is associated with expansion of IL-17A Producing gamma Delta T cells. Frontiers in immunology 8(2017).
61. Rooks, M.G., et al. Gut microbiome composition and function in experimental colitis during active disease and treatment-induced remission. Isme J 8, 1403-1417 (2014).
62. Simpson, K.W., et al. Adherent and invasive Escherichia coli is associated with granulomatous colitis in boxer dogs. Infection and immunity 74, 4778-4792 (2006).
63. Janeczko, S., et al. The relationship of mucosal bacteria to duodenal histopathology, cytokine mRNA, and clinical disease activity in cats with inflammatory bowel disease. Veterinary microbiology 128, 178-193 (2008).
64. Larmonier, C.B., et al. Reduced colonic microbial diversity is associated with colitis in NHE3-deficient mice. Am J Physiol Gastrointest Liver Physiol 305, G667-677 (2013).

65. Spalinger, M.R., et al. PTPN2 controls differentiation of CD4(+) T cells and limits intestinal inflammation and intestinal dysbiosis. Mucosal Immunol 8, 918-929 (2015).
66. Munyaka, P.M., Rabbi, M.F., Khafipour, E. & Ghia, J.E. Acute dextran sulfate sodium (DSS)-induced colitis promotes gut microbial dysbiosis in mice. J Basic Microbiol 56, 986-998 (2016).
67. Okayasu, I., et al. A novel method in the induction of reliable experimental acute and chronic ulcerative colitis in mice. Gastroenterology 98, 694-702 (1990).
68. Robinson, A.M., et al. Alterations of colonic function in the Winnie mouse model of spontaneous chronic colitis. Am J Physiol Gastrointest Liver Physiol 312, G85-G102 (2017).
69. Yang, I., et al. Intestinal microbiota composition of interleukin-10 deficient C57BL/6J mice and susceptibility to Helicobacter hepaticus-induced colitis. PloS one 8, e70783 (2013).
70. Ren, Y., et al. Polysaccharide of Hericium erinaceus attenuates colitis in C57BL/6 mice via regulation of oxidative stress, inflammation-related signaling pathways and modulating the composition of the gut microbiota. J Nutr Biochem 57, 67-76 (2018).
71. Nunberg, M., et al. Interleukin 1alpha-Deficient Mice Have an Altered Gut Microbiota Leading to Protection from Dextran Sodium Sulfate-Induced Colitis. mSystems 3(2018).
72. Schaubeck, M., et al. Dysbiotic gut microbiota causes transmissible Crohn's disease-like ileitis independent of failure in antimicrobial defence. Gut 65, 225-237 (2016).
73. Lupp, C., et al. Host-mediated inflammation disrupts the intestinal microbiota and promotes the overgrowth of Enterobacteriaceae. Cell host & microbe 2, 204 (2007).
74. Nones, K., et al. Multidrug resistance gene deficient (mdr1a-/-) mice have an altered caecal microbiota that precedes the onset of intestinal inflammation. J Appl Microbiol 107, 557-566 (2009).
75. Manchester, A.C., et al. Association between Granulomatous Colitis in French Bulldogs and Invasive Escherichia coli and Response to Fluoroquinolone Antimicrobials. J Vet Intern Med 27, 56-61 (2013).
76. Roy, U., et al. Distinct Microbial Communities Trigger Colitis Development upon Intestinal Barrier Damage via Innate or Adaptive Immune Cells. Cell Rep 21, 994-1008 (2017).
77. Johnston, D.G.W., et al. Loss of microRNA-21 Influences the Gut Microbiota Causing Reduced Susceptibility in a Murine Model of Colitis. J Crohns Colitis (2018).
78. Berry, D., et al. Phylotype-level 16S rRNA analysis reveals new bacterial indicators of health state in acute murine colitis. Isme J 6, 2091-2106 (2012).
79. Zhang, Q., et al. Accelerated dysbiosis of gut microbiota during aggravation of DSS-induced colitis by a butyrate-producing bacterium. Sci Rep 6, 27572 (2016).
80. Schwab, C., et al. Longitudinal study of murine microbiota activity and interactions with the host during acute inflammation and recovery. Isme J 8, 1101-1114 (2014).
81. Zimmermann, J., et al. The intestinal microbiota determines the colitis-inducing potential of T-bet-deficient Th cells in mice. Eur J Immunol 48, 161-167 (2018).
82. Garrett, W.S., et al. Enterobacteriaceae act in concert with the gut microbiota to induce spontaneous and maternally transmitted colitis. Cell host & microbe 8, 292-300 (2010).
83. Palm, N.W., et al. Immunoglobulin A coating identifies colitogenic bacteria in inflammatory bowel disease. Cell 158, 1000-1010 (2014).
84. Maharshak, N., et al. Altered enteric microbiota ecology in interleukin 10-deficient mice during development and progression of intestinal inflammation. Gut microbes 4, 316-324 (2013).
85. Chassaing, B., et al. Dietary emulsifiers impact the mouse gut microbiota promoting colitis and metabolic syndrome. Nature 519, 92-96 (2015).
86. Osaka, T., et al. Meta-Analysis of Fecal Microbiota and Metabolites in Experimental Colitic Mice during the Inflammatory and Healing Phases. Nutrients 9(2017).
87. Perez-Munoz, M.E., et al. Discordance between changes in the gut microbiota and pathogenicity in a mouse model of spontaneous colitis. Gut microbes 5, 286-295 (2014).
88. Nagalingam, N.A., Kao, J.Y. & Young, V.B. Microbial ecology of the murine gut associated with the development of dextran sodium sulfate-induced colitis. Inflamm Bowel Dis 17, 917-926 (2011).
89. Alkadhi, S., Kunde, D., Cheluvappa, R., Randall-Demllo, S. & Eri, R. The murine appendiceal microbiome is altered in spontaneous colitis and its pathological progression. Gut Pathog 6, 25 (2014).
90. Carvalho, F.A., et al. Interleukin-1beta (IL-1beta) promotes susceptibility of Toll-like receptor 5 (TLR5) deficient mice to colitis. Gut 61, 373-384 (2012).

91. Seregin, S.S., et al. NLRP6 Protects Il10(-/-) Mice from Colitis by Limiting Colonization of Akkermansia muciniphila. Cell Rep 19, 2174 (2017).
92. Selvanantham, T., et al. NKT Cell-Deficient Mice Harbor an Altered Microbiota That Fuels Intestinal Inflammation during Chemically Induced Colitis. J Immunol 197, 4464-4472 (2016).
93. Lamas, B., et al. CARD9 impacts colitis by altering gut microbiota metabolism of tryptophan into aryl hydrocarbon receptor ligands. Nature medicine 22, 598-+ (2016).
94. Madsen, K.L., et al. Antibiotic therapy attenuates colitis in interleukin 10 gene-deficient mice. Gastroenterology 118, 1094-1105 (2000).
95. Mitchell, J., et al. Colonic Inhibition of Phosphatase and Tensin Homolog Increases Colitogenic Bacteria, Causing Development of Colitis in Il10-/- Mice. Inflamm Bowel Dis (2018).
96. Hakansson, A., et al. Immunological alteration and changes of gut microbiota after dextran sulfate sodium (DSS) administration in mice. Clin Exp Med 15, 107-120 (2015).
97. Heimesaat, M.M., et al. Shift towards pro-inflammatory intestinal bacteria aggravates acute murine colitis via Toll-like receptors 2 and 4. PloS one 2, e662 (2007).
98. Wohlgemuth, S., Haller, D., Blaut, M. & Loh, G. Reduced microbial diversity and high numbers of one single Escherichia coli strain in the intestine of colitic mice. Environ Microbiol 11, 1562-1571 (2009).
99. Arthur, J.C., et al. Intestinal inflammation targets cancer-inducing activity of the microbiota. Science 338, 120-123 (2012).
100. Schuppler, M., Lotzsch, K., Waidmann, M. & Autenrieth, I.B. An abundance of Escherichia coli is harbored by the mucosa-associated bacterial flora of interleukin-2-deficient mice. Infection and immunity 72, 1983-1990 (2004).
101. Hoentjen, F., et al. Antibiotics with a selective aerobic or anaerobic spectrum have different therapeutic activities in various regions of the colon in interleukin 10 gene deficient mice. Gut 52, 1721-1727 (2003).
102. Zenewicz, L.A., et al. IL-22 deficiency alters colonic microbiota to be transmissible and colitogenic. J Immunol 190, 5306-5312 (2013).
103. Dennis, K.L., et al. Adenomatous polyps are driven by microbe-instigated focal inflammation and are controlled by IL-10-producing T cells. Cancer Res 73, 5905-5913 (2013).
104. Abecia, L.H., L.; Khoo, C.; Frantz, N.; McCartney, A.;. Effects of a novel galactooligosaccharide on the faecal microbiota of healthy and inflammatory bowel disease International Journal of Probiotics and Prebiotics 5, 61-68 (2010).
105. Xu, J., et al. Does canine inflammatory bowel disease influence gut microbial profile and host metabolism? Bmc Vet Res 12(2016).
106. Jones-Hall, Y.L., Kozik, A. & Nakatsu, C. Ablation of tumor necrosis factor is associated with decreased inflammation and alterations of the microbiota in a mouse model of inflammatory bowel disease. PloS one 10, e0119441 (2015).
107. Bel, S., et al. Reprogrammed and transmissible intestinal microbiota confer diminished susceptibility to induced colitis in TMF-/- mice. Proceedings of the National Academy of Sciences of the United States of America 111, 4964-4969 (2014).
108. Bloom, S.M., et al. Commensal Bacteroides species induce colitis in host-genotype-specific fashion in a mouse model of inflammatory bowel disease. Cell host & microbe 9, 390-403 (2011).
109. He, Q., et al. Dysbiosis of the fecal microbiota in the TNBS-induced Crohn's disease mouse model. Appl Microbiol Biotechnol 100, 4485-4494 (2016).
110. Couturier-Maillard, A., et al. NOD2-mediated dysbiosis predisposes mice to transmissible colitis and colorectal cancer. The Journal of clinical investigation 123, 700-711 (2013).
111. Ettreiki, C., et al. Juvenile ferric iron prevents microbiota dysbiosis and colitis in adult rodents. World journal of gastroenterology : WJG 18, 2619-2629 (2012).
112. Devkota, S., et al. Dietary-fat-induced taurocholic acid promotes pathobiont expansion and colitis in Il10-/- mice. Nature 487, 104-108 (2012).
113. Zhang, Z., et al. Chlorogenic Acid Ameliorates Experimental Colitis by Promoting Growth of Akkermansia in Mice. Nutrients 9(2017).
114. Ghosh, S., Molcan, E., DeCoffe, D., Dai, C. & Gibson, D.L. Diets rich in n-6 PUFA induce intestinal microbial dysbiosis in aged mice. Br J Nutr 110, 515-523 (2013).

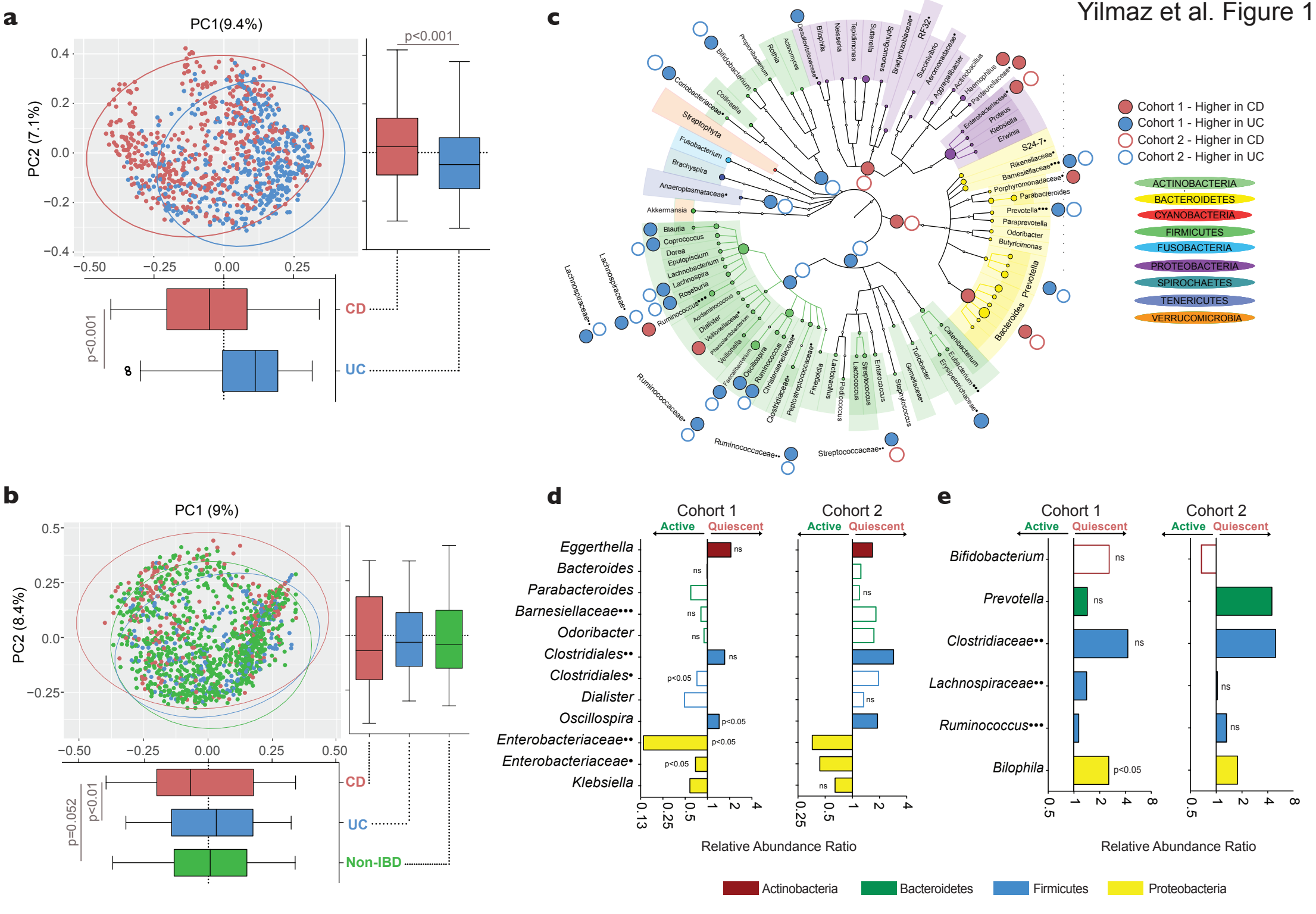
115. Ye, J., et al. Bacteria and bacterial rRNA genes associated with the development of colitis in IL-10(-/-) mice. Inflamm Bowel Dis 14, 1041-1050 (2008).
116. Vijay-Kumar, M., et al. Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5. Science 328, 228-231 (2010).
117. Zhu, W., et al. Precision editing of the gut microbiota ameliorates colitis. Nature 553, 208-211 (2018).
118. Moon, C., Stupp, G.S., Su, A.I. & Wolan, D.W. Metaproteomics of Colonic Microbiota Unveils Discrete Protein Functions among Colitic Mice and Control Groups. Proteomics 18(2018).
119. Du, Z., et al. Development of gut inflammation in mice colonized with mucosa-associated bacteria from patients with ulcerative colitis. Gut Pathog 7, 32 (2015).
120. Sokol, H., et al. Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. Proceedings of the National Academy of Sciences of the United States of America 105, 16731-16736 (2008).
121. Bibiloni, R., Simon, M.A., Albright, C., Sartor, B. & Tannock, G.W. Analysis of the large bowel microbiota of colitic mice using PCR/DGGE. Lett Appl Microbiol 41, 45-51 (2005).
122. Kudelka, M.R., et al. Cosmc is an X-linked inflammatory bowel disease risk gene that spatially regulates gut microbiota and contributes to sex-specific risk. Proceedings of the National Academy of Sciences of the United States of America 113, 14787-14792 (2016).
123. Vereecke, L., et al. A20 controls intestinal homeostasis through cell-specific activities. Nat Commun 5, 5103 (2014).
124. Pittet, V*., et al.* Cohort profile: the Swiss Inflammatory Bowel Disease Cohort Study (SIBDCS). *Int J Epidemiol* **38**, 922-931 (2009).
125. Lennard-Jones, J.E. Classification of inflammatory bowel disease. *Scand J Gastroenterol Suppl* **170**, 2-6; discussion 16-19 (1989).
126. Harris, P.A*., et al.* Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* **42**, 377-381 (2009).
127. Gevers, D*., et al.* The treatment-naive microbiome in new-onset Crohn's disease. *Cell host & microbe* **15**, 382-392 (2014).
128. Sundquist, A*., et al.* Bacterial flora-typing with targeted, chip-based Pyrosequencing. *Bmc Microbiol* **7**(2007).
129. Yilmaz, B*., et al.* The presence of genetic risk variants within PTPN2 and PTPN22 is associated with intestinal microbiota alterations in Swiss IBD cohort patients. *PloS one* **13**, e0199664 (2018).
130. Whiteley, A.S*., et al.* Microbial 16S rRNA Ion Tag and community metagenome sequencing using the Ion Torrent (PGM) Platform. *J Microbiol Methods* **91**, 80-88 (2012).
131. Caporaso, J.G*., et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**, 335-336 (2010).
132. McMurdie, P.J. & Holmes, S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PloS one* **8**(2013).
133. Callahan, B.J., Sankaran, K., Fukuyama, J.A., McMurdie, P.J. & Holmes, S.P. Bioconductor Workflow for Microbiome Data Analysis: from raw reads to community analyses. *F1000Res* **5**, 1492 (2016).
134. Rahnavard, G.F., Eric, A.; McIver, Lauren, J.; Schwager, Emma; Weingart, George, Moon, Yo Sup ;Morgan, Xochitl C. ; Waldron, Levi; Huttenhower, Curtis High-sensitivity pattern discovery in large multi'omic datasets. (2017).
135. Morgan, X.C*., et al.* Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol* **13**, R79 (2012).
136. Edgar, R.C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460-2461 (2010).
137. Good, I.J. The population frequencies of species and the estimation of population parameters. *Biometrika* **40**, 237-264 (1953).
138. Williams, R.J., Howe, A. & Hofmockel, K.S. Demonstrating microbial co-occurrence pattern analyses within and between ecosystems. *Frontiers in Microbiology* **5**(2014).
139. Su, G., Kuchinsky, A., Morris, J.H., States, D.J. & Meng, F. GLay: community structure analysis of biological networks. *Bioinformatics* **26**, 3135-3137 (2010).

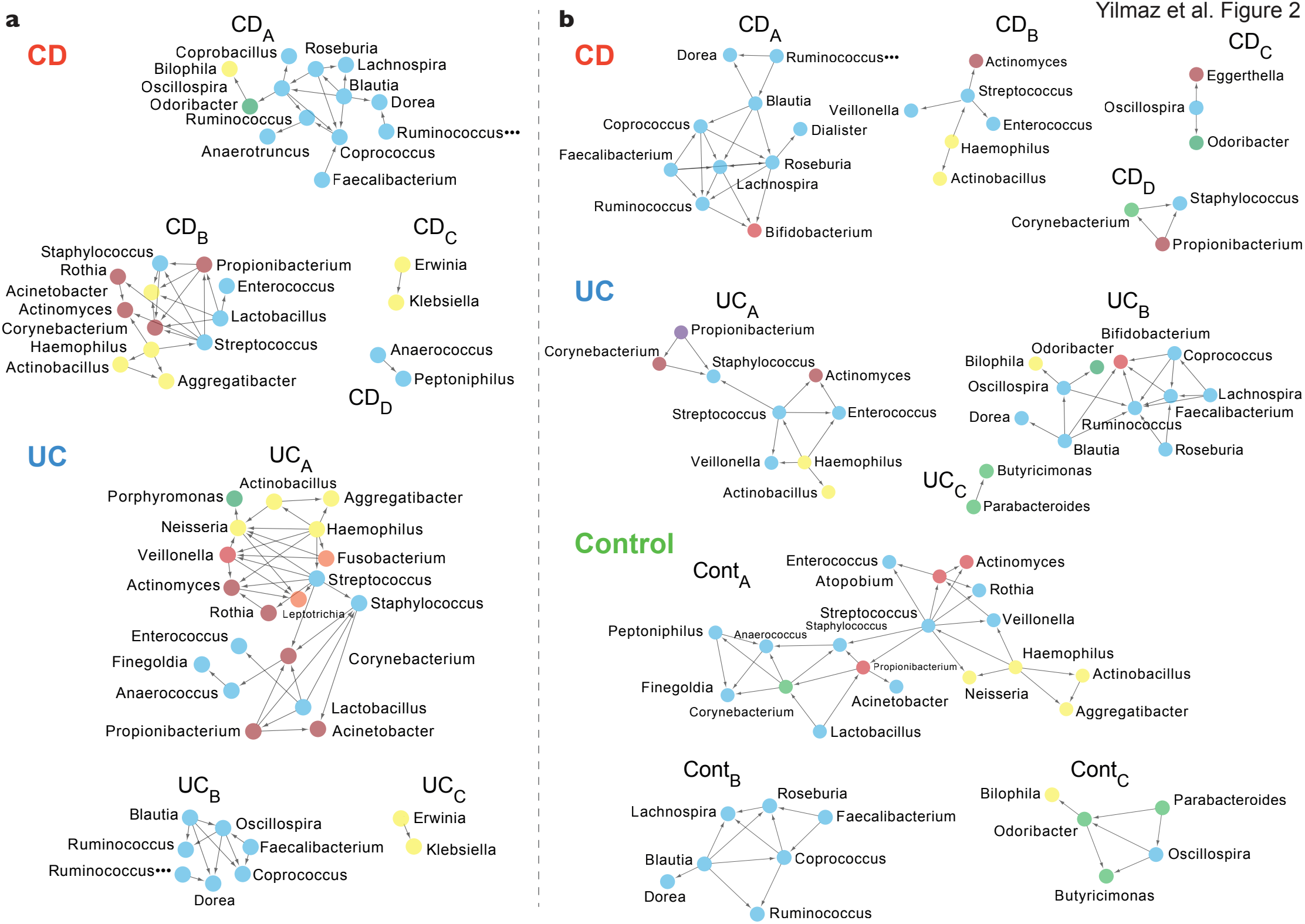
140. Magnusdottir, S.*, et al.* Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nat Biotechnol* **35**, 81-89 (2017).
141. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *J Roy Stat Soc B Met* **57**, 289-300 (1995).

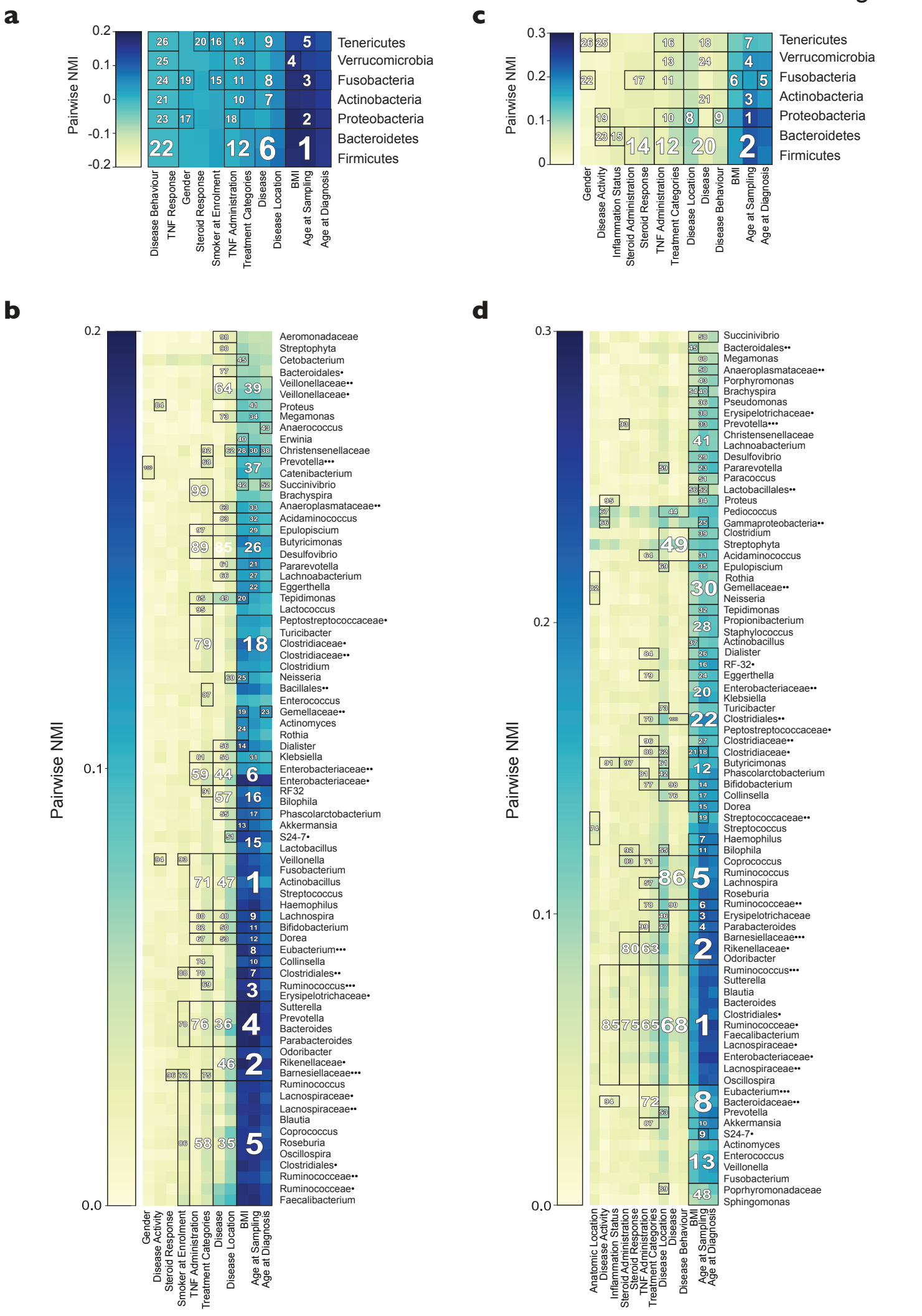
Data Availability Statement

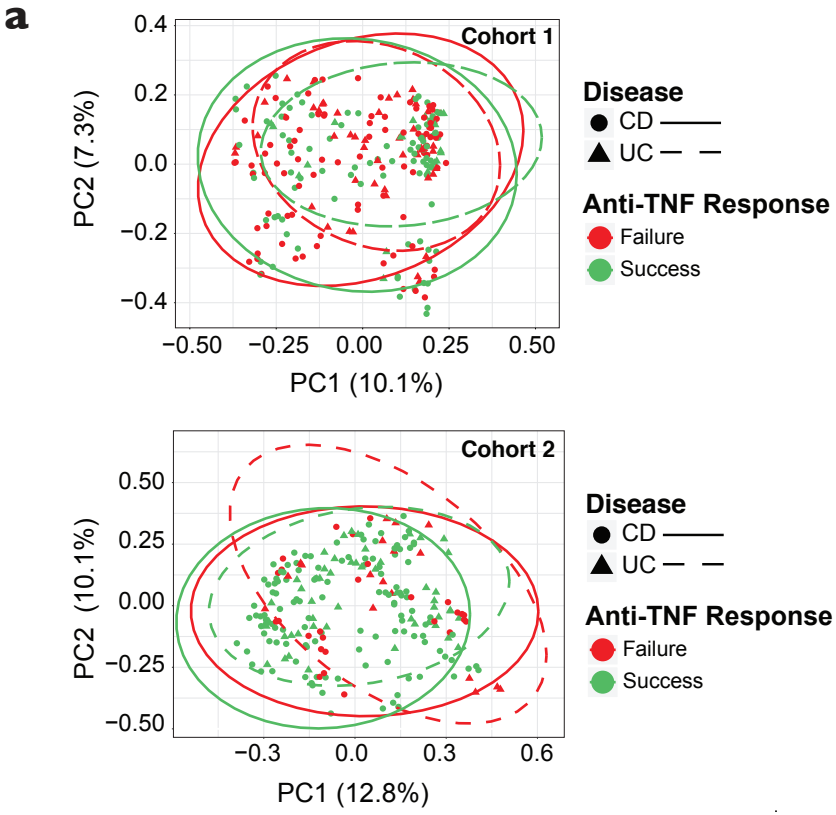
All sequencing datasets from the current study have been deposited in figshare repository and are publicly available. Cohort 1 fasta file with mapping file are available at https://figshare.com/s/e9f2cffd0f0328ca5811 (DOI: 10.6084/m9.figshare.7335068) and Cohort 2 fasta file with mapping file are available at https://figshare.com/s/bbdd5dfb01e29484efa1 (DOI: 10.6084/m9.figshare.7335071). Associated codes for the analysis using R packages and QIIME can be found in these depositories.

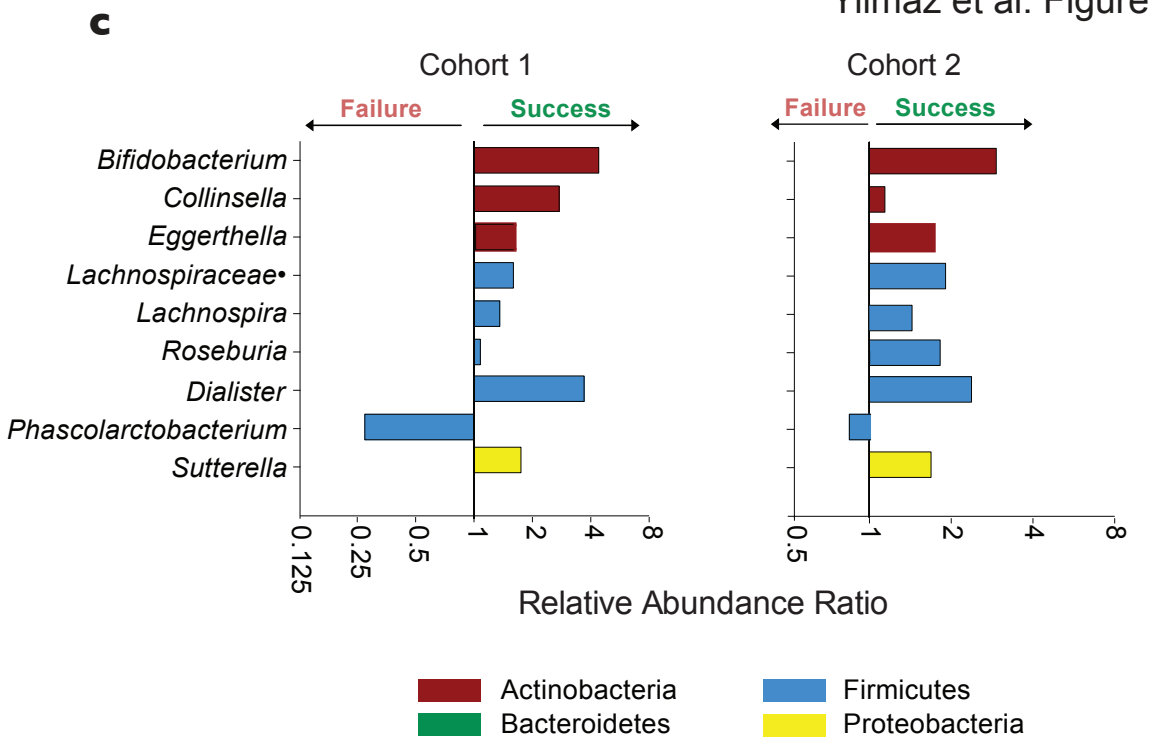
Yilmaz et al. Figure 1

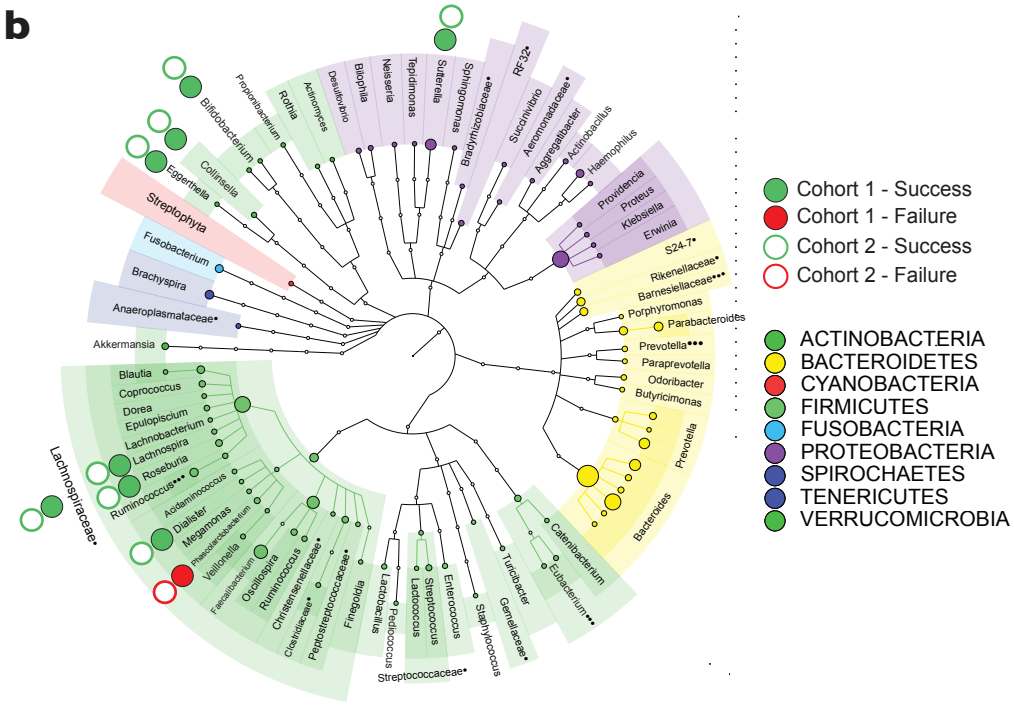


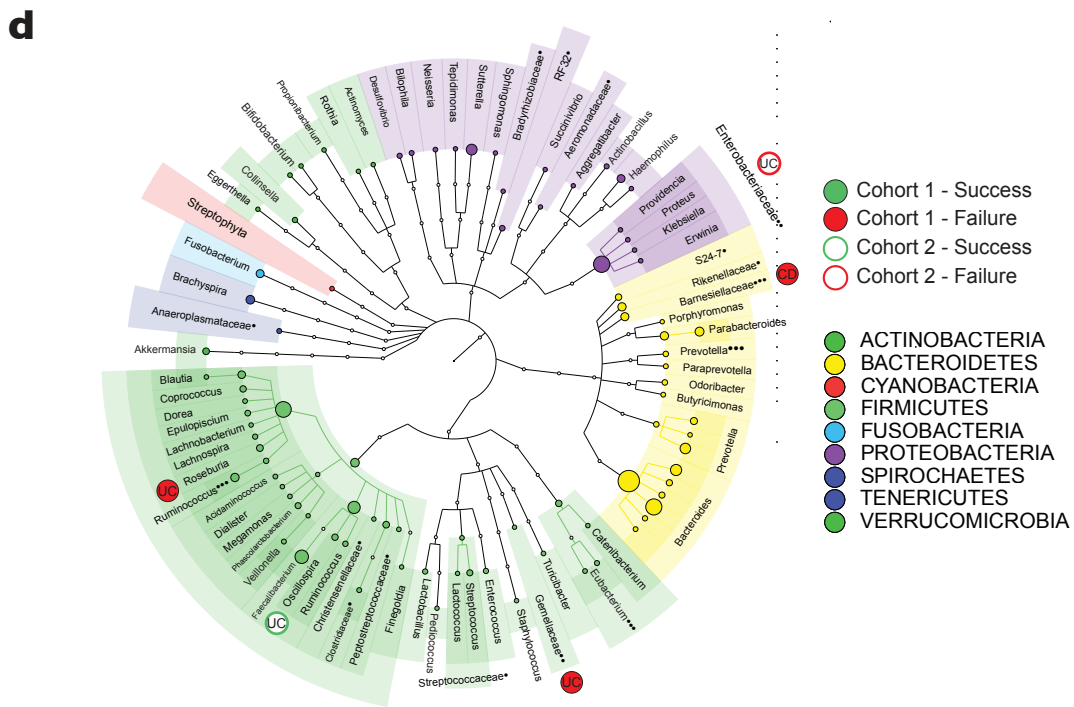

a
CD
CD_A
Coprobacillus
Roseburia
Bilophila
Lachnospira
Oscillospira
Blautia
Odoribacter
Dorea
Ruminococcus
Ruminococcus•••
Anaerotruncus
Coprococcus
Faecalibacterium
CD_B
Staphylococcus
Propionibacterium
Rothia
Enterococcus
Acinetobacter
Actinomyces
Lactobacillus
Corynebacterium
Streptococcus
Haemophilus
Actinobacillus
Aggregatibacter
CD_C
Erwinia
Klebsiella
Anaerococcus
Peptoniphilus
CD_D
UC
UC_A
Actinobacillus
Porphyromonas
Aggregatibacter
Neisseria
Haemophilus
Veillonella
Fusobacterium
Actinomyces
Streptococcus
Rothia
Leptotrichia
Staphylococcus
Enterococcus
Finegoldia
Corynebacterium
Anaerococcus
Lactobacillus
Propionibacterium
Acinetobacter
UC_B
Blautia
Oscillospira
Ruminococcus
Faecalibacterium
Ruminococcus•••
Coprococcus
Dorea
UC_C
Erwinia
Klebsiella
b
CD
CD_A
Dorea
Ruminococcus•••
Blautia
Coprococcus
Dialister
Faecalibacterium
Roseburia
Lachnospira
Ruminococcus
Bifidobacterium
CD_B
Actinomyces
Streptococcus
Veillonella
Enterococcus
Haemophilus
Actinobacillus
CD_C
Eggerthella
Oscillospira
Odoribacter
CD_D
Staphylococcus
Corynebacterium
Propionibacterium
UC
UC_A
Propionibacterium
Corynebacterium
Staphylococcus
Actinomyces
Streptococcus
Enterococcus
Veillonella
Haemophilus
Actinobacillus
UC_B
Bifidobacterium
Odoribacter
Coprococcus
Bilophila
Oscillospira
Lachnospira
Faecalibacterium
Dorea
Ruminococcus
Blautia
Roseburia
UC_C
Butyricimonas
Parabacteroides
Control
Cont_A
Enterococcus
Actinomyces
Atopobium
Rothia
Streptococcus
Staphylococcus
Veillonella
Peptoniphilus
Anaerococcus
Haemophilus
Propionibacterium
Actinobacillus
Finegoldia
Neisseria
Acinetobacter
Corynebacterium
Aggregatibacter
Lactobacillus
Cont_B
Roseburia
Lachnospira
Faecalibacterium
Blautia
Coprococcus
Dorea
Ruminococcus
Cont_C
Bilophila
Parabacteroides
Odoribacter
Oscillospira
Butyricimonas

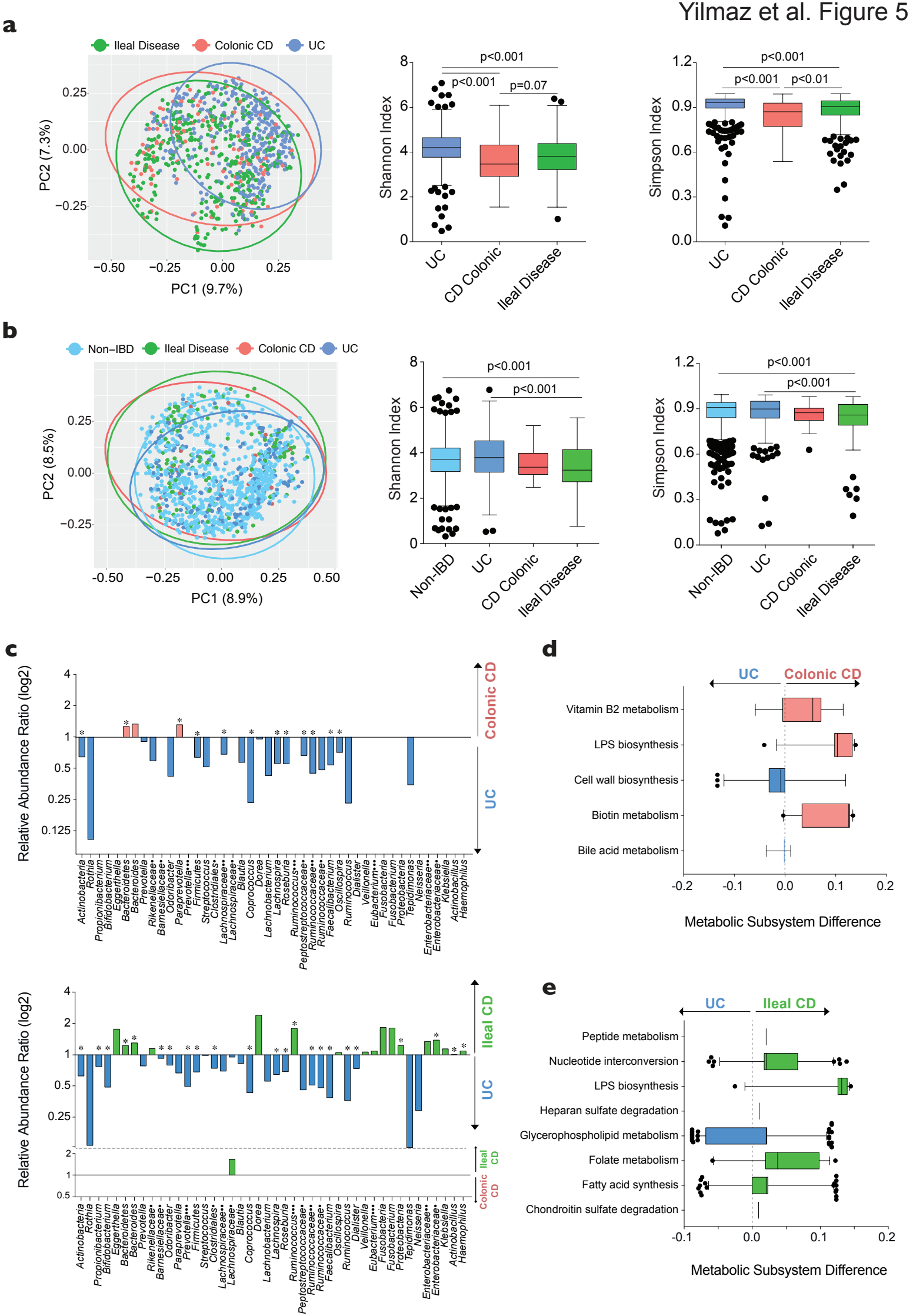

a
Pairwise NMI
0.2
0.1
0
-0.1
-0.2
Tenericutes
Verrucomicrobia
Fusobacteria
Actinobacteria
Proteobacteria
Bacteroidetes
Firmicutes
Disease Behaviour
TNF Response
Gender
Steroid Response
Smoker at Enrolment
TNF Administration
Treatment Categories
Disease
Disease Location
BMI
Age at Sampling
Age at Diagnosis
c
Pairwise NMI
0.3
0.2
0.1
0
Tenericutes
Verrucomicrobia
Fusobacteria
Actinobacteria
Proteobacteria
Bacteroidetes
Firmicutes
Gender
Disease Activity
Inflammation Status
Steroid Administration
Steroid Response
TNF Administration
Treatment Categories
Disease Location
Disease
Disease Behaviour
BMI
Age at Sampling
Age at Diagnosis
b
Pairwise NMI
0.2
0.1
0.0
Aeromonadaceae
Streptophyta
Cetobacterium
Bacteroidales•
Veillonellaceae••
Veillonellaceae•
Proteus
Megamonas
Anaerococcus
Erwinia
Christensenellaceae
Prevotella•••
Catenibacterium
Succinivibrio
Brachyspira
Anaeroplasmataceae••
Acidaminococcus
Epulopiscium
Butyricimonas
Desulfovibrio
Pararevotella
Lachnoabacterium
Eggerthella
Tepidimonas
Lactococcus
Peptostreptococcaceae•
Turicibacter
Clostridiaceae•
Clostridiaceae••
Clostridium
Neisseria
Bacillales••
Enterococcus
Gemellaceae••
Actinomyces
Rothia
Dialister
Klebsiella
Enterobacteriaceae••
Enterobacteriaceae•
RF32
Bilophila
Phascolarctobacterium
Akkermansia
S24-7•
Lactobacillus
Veillonella
Fusobacterium
Actinobacillus
Streptococcus
Haemophilus
Lachnospira
Bifidobacterium
Dorea
Eubacterium•••
Collinsella
Clostridiales••
Ruminococcus•••
Erysipelotrichaceae•
Sutterella
Prevotella
Bacteroides
Parabacteroides
Odoribacter
Rikenellaceae•
Barnesiellaceae•••
Ruminococcus
Lacnospiraceae•
Lacnospiraceae••
Blautia
Coprococcus
Roseburia
Oscillospira
Clostridiales•
Ruminococceae••
Ruminococceae•
Faecalibacterium
Gender
Disease Activity
Steroid Response
Smoker at Enrolment
TNF Administration
Treatment Categories
Disease
Disease Location
BMI
Age at Sampling
Age at Diagnosis
d
Pairwise NMI
0.3
0.2
0.1
0.0
Succinivibrio
Bacteroidales••
Megamonas
Anaeroplasmataceae••
Porphyromonas
Brachyspira
Pseudomonas
Erysipelotrichaceae•
Prevotella•••
Christensenellaceae
Lachnoabacterium
Desulfovibrio
Pararevotella
Paracoccus
Lactobacillales••
Proteus
Pediococcus
Gammaproteobacteria••
Clostridium
Streptophyta
Acidaminococcus
Epulopiscium
Rothia
Gemellaceae••
Neisseria
Tepidimonas
Propionibacterium
Staphylococcus
Actinobacillus
Dialister
RF-32•
Eggerthella
Enterobacteriaceae••
Klebsiella
Turicibacter
Clostridiales••
Peptostreptococcaceae•
Clostridiaceae••
Clostridiaceae•
Butyricimonas
Phascolarctobacterium
Bifidobacterium
Collinsella
Dorea
Streptococcaceae••
Streptococcus
Haemophilus
Bilophila
Coprococcus
Ruminococcus
Lachnospira
Roseburia
Ruminococceae••
Erysipelotrichaceae
Parabacteroides
Barnesiellaceae•••
Rikenellaceae•
Odoribacter
Ruminococcus•••
Sutterella
Blautia
Bacteroides
Clostridiales•
Ruminococceae•
Faecalibacterium
Lacnospiraceae•
Enterobacteriaceae•
Lacnospiraceae••
Oscillospira
Eubacterium•••
Bacteroidaceae••
Prevotella
Akkermansia
S24-7•
Actinomyces
Enterococcus
Veillonella
Fusobacterium
Poprhyromonadaceae
Sphingomonas
Anatomic Location
Disease Activity
Inflammation Status
Steroid Administration
Steroid Response
TNF Administration
Treatment Categories
Disease Location
Disease
Disease Behaviour
BMI
Age at Sampling
Age at Diagnosis

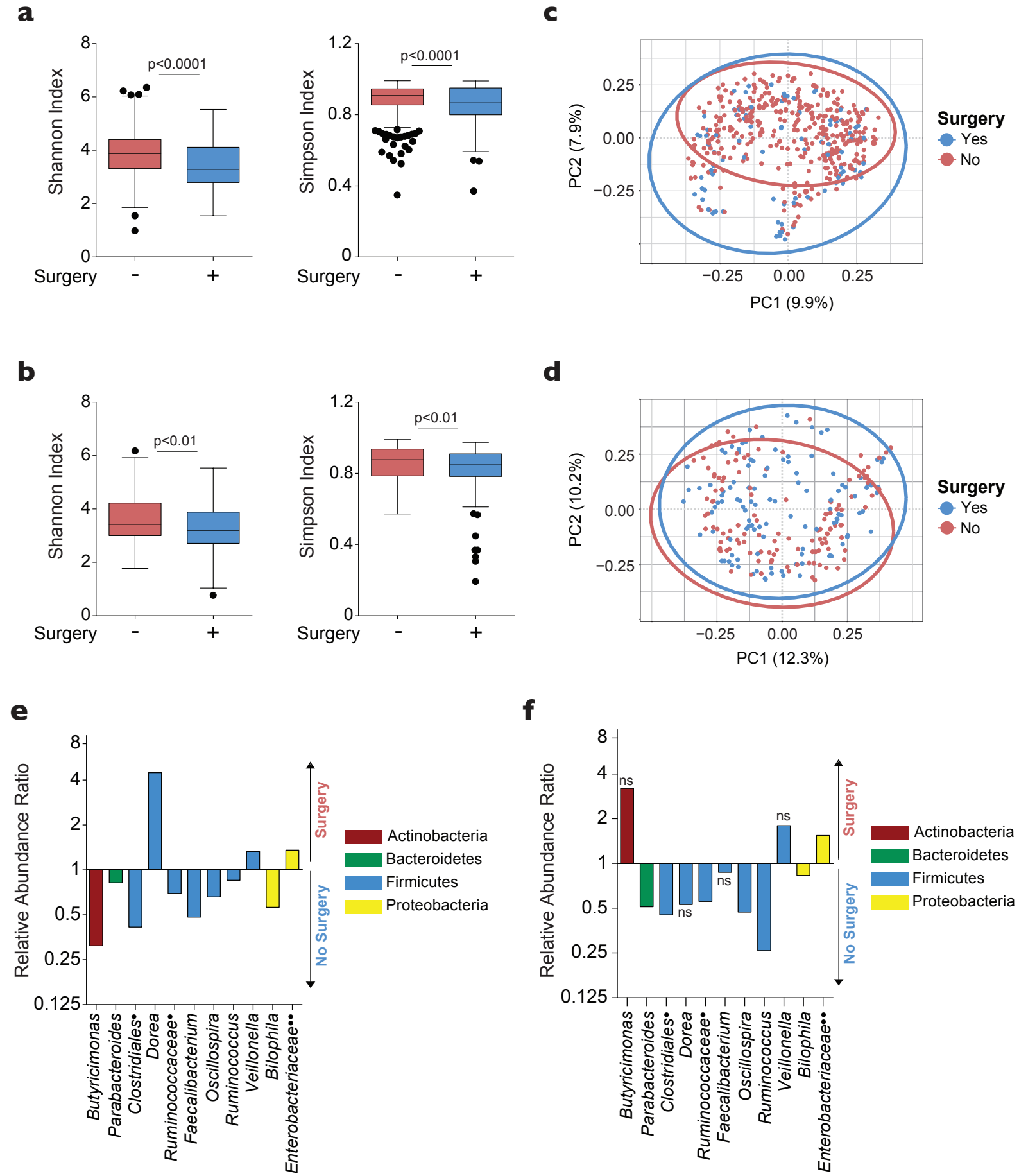

a
Cohort 1
PC2 (7.3%)
PC1 (10.1%)
Disease
CD
UC
Anti-TNF Response
Failure
Success
Cohort 2
PC2 (10.1%)
PC1 (12.8%)
Disease
CD
UC
Anti-TNF Response
Failure
Success


c
Cohort 1
Cohort 2
Failure
Success
Bifidobacterium
Collinsella
Eggerthella
Lachnospiraceae•
Lachnospira
Roseburia
Dialister
Phascolarctobacterium
Sutterella
Relative Abundance Ratio
Actinobacteria
Bacteroidetes
Firmicutes
Proteobacteria

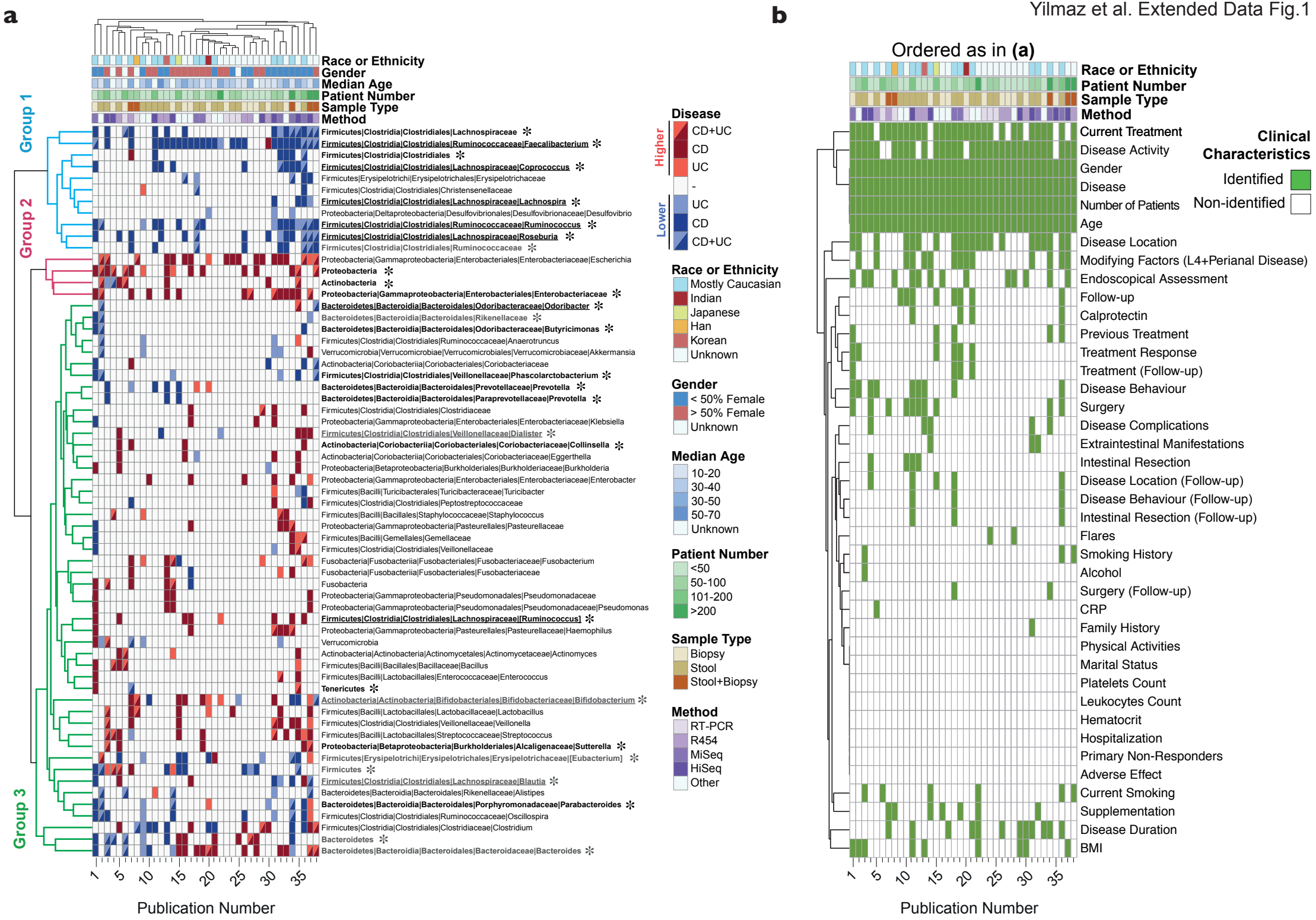

b
Cohort 1 - Success
Cohort 1 - Failure
Cohort 2 - Success
Cohort 2 - Failure
ACTINOBACTERIA
BACTEROIDETES
CYANOBACTERIA
FIRMICUTES
FUSOBACTERIA
PROTEOBACTERIA
SPIROCHAETES
TENERICUTES
VERRUCOMICROBIA


d
Cohort 1 - Success
Cohort 1 - Failure
Cohort 2 - Success
Cohort 2 - Failure
ACTINOBACTERIA
BACTEROIDETES
CYANOBACTERIA
FIRMICUTES
FUSOBACTERIA
PROTEOBACTERIA
SPIROCHAETES
TENERICUTES
VERRUCOMICROBIA

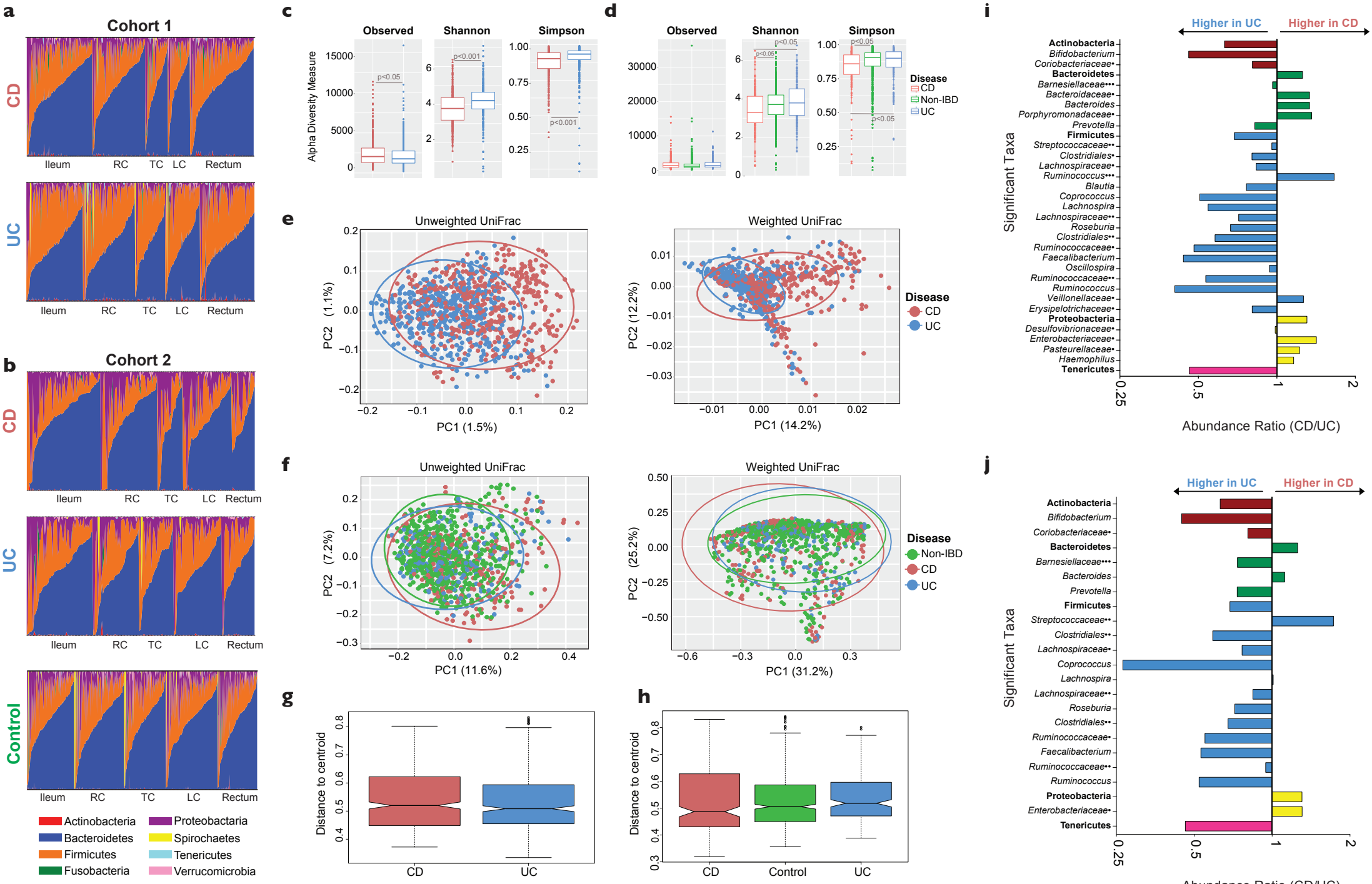

a
Ileal Disease
Colonic CD
UC
PC2 (7.3%)
PC1 (9.7%)
Shannon Index
p<0.001
p<0.001
p=0.07
UC
CD Colonic
Ileal Disease
Simpson Index
p<0.001
p<0.001
p<0.01
b
Non-IBD
Ileal Disease
Colonic CD
UC
PC2 (8.5%)
PC1 (8.9%)
Shannon Index
p<0.001
p<0.001
Non-IBD
UC
CD Colonic
Ileal Disease
Simpson Index
p<0.001
p<0.001
c
Relative Abundance Ratio (log2)
Colonic CD
UC
Ileal CD
d
UC
Colonic CD
Vitamin B2 metabolism
LPS biosynthesis
Cell wall biosynthesis
Biotin metabolism
Bile acid metabolism
Metabolic Subsystem Difference
e
UC
Ileal CD
Peptide metabolism
Nucleotide interconversion
LPS biosynthesis
Heparan sulfate degradation
Glycerophospholipid metabolism
Folate metabolism
Fatty acid synthesis
Chondroitin sulfate degradation
Metabolic Subsystem Difference

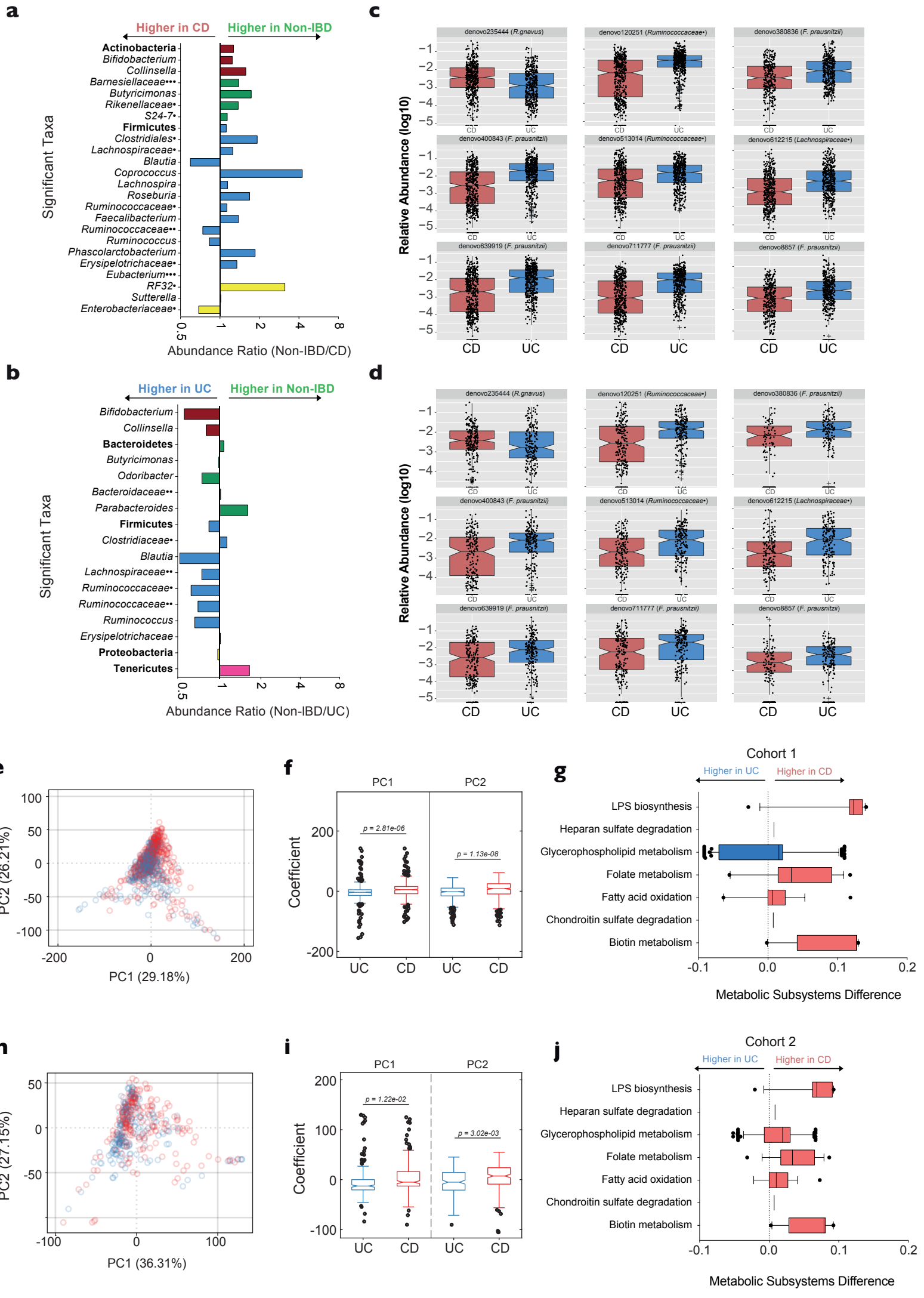

a
Shannon Index
p<0.0001
Surgery
Simpson Index
p<0.0001
b
Shannon Index
p<0.01
Simpson Index
p<0.01
c
PC2 (7.9%)
PC1 (9.9%)
Surgery
Yes
No
d
PC2 (10.2%)
PC1 (12.3%)
e
Relative Abundance Ratio
Surgery
No Surgery
Actinobacteria
Bacteroidetes
Firmicutes
Proteobacteria
Butyricimonas
Parabacteroides
Clostridiales•
Dorea
Ruminococcaceae•
Faecalibacterium
Oscillospira
Ruminococcus
Veillonella
Bilophila
Enterobacteriaceae••
f
ns

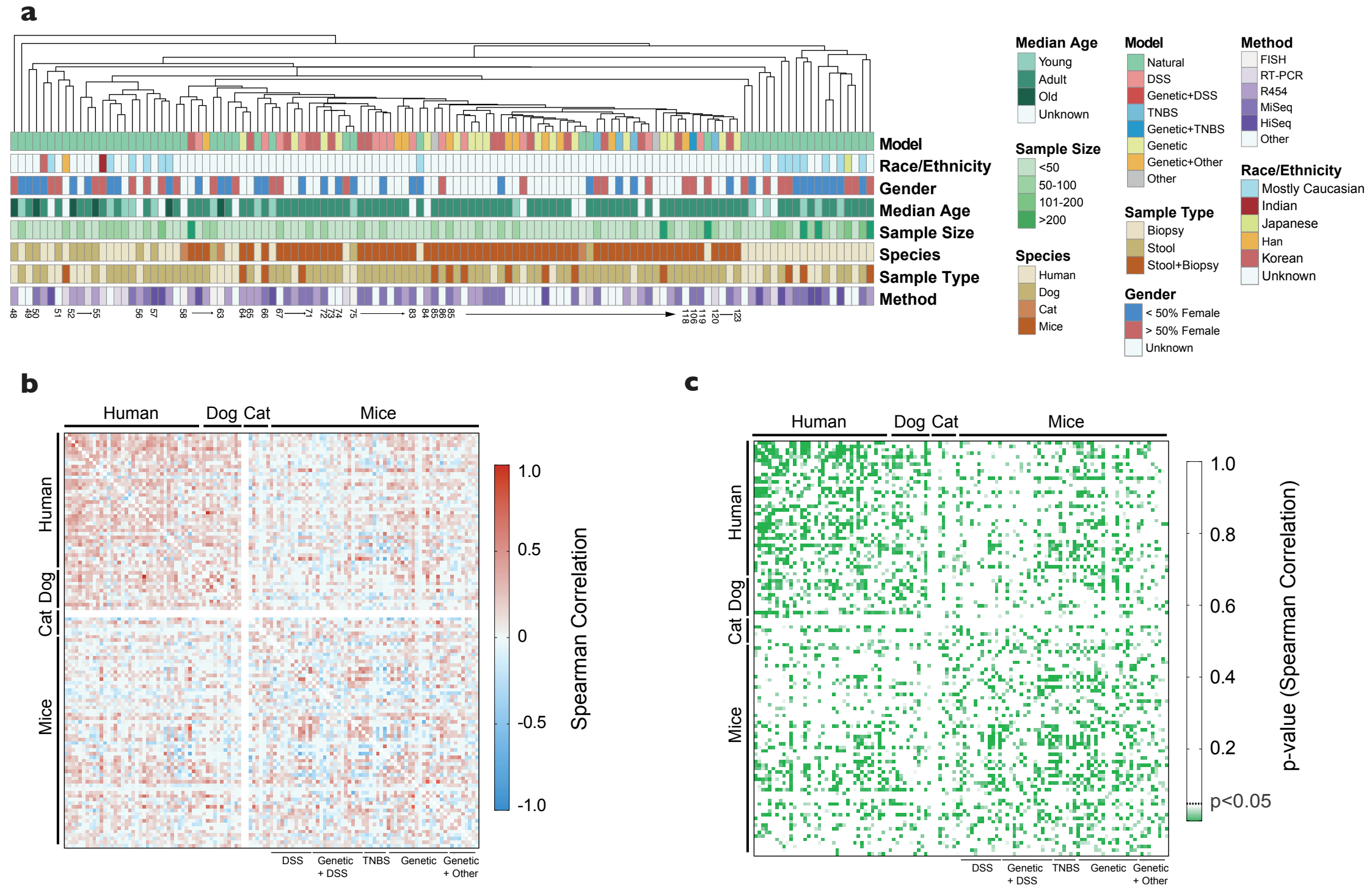
Yilmaz et al. Extended Data Fig.1

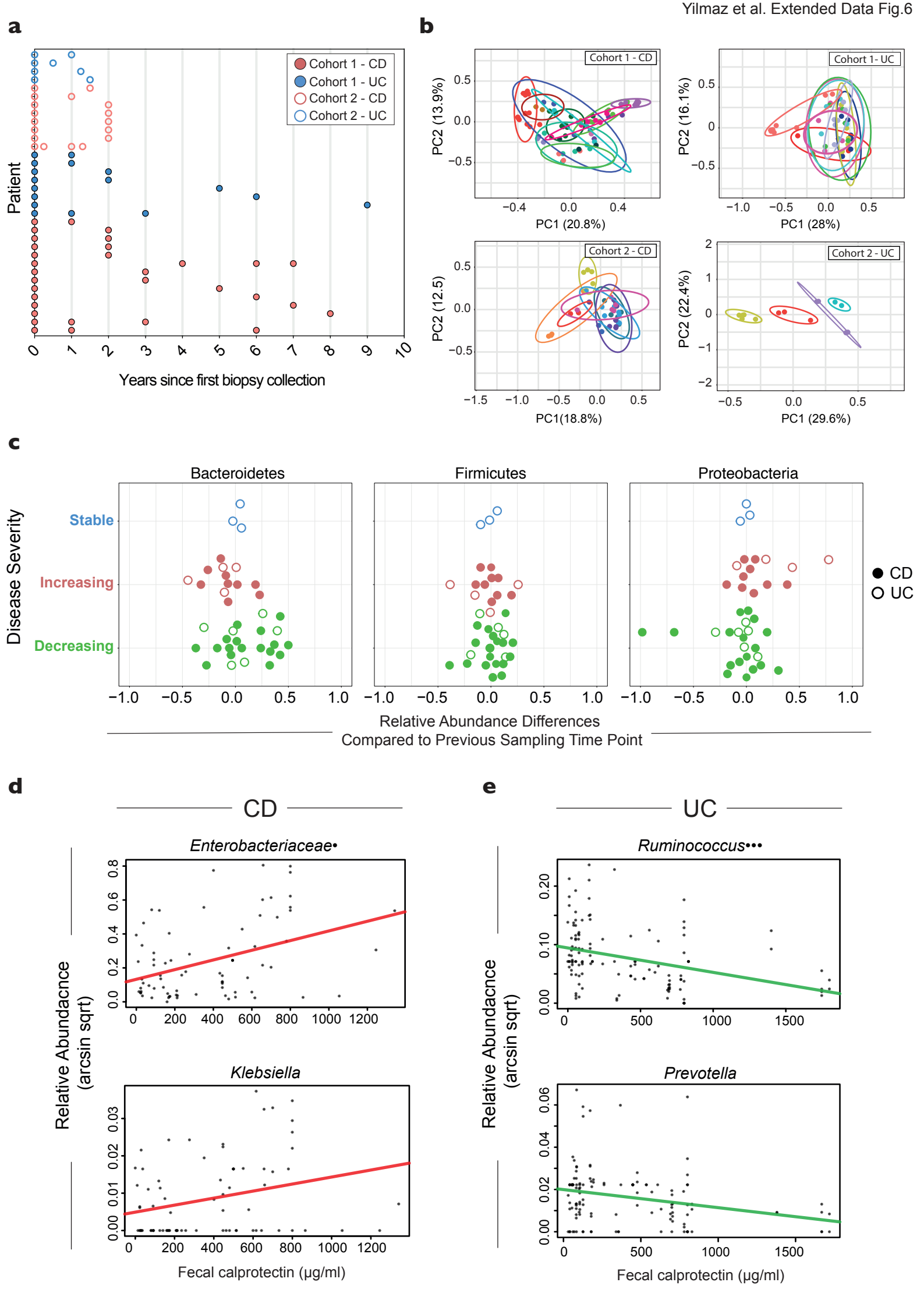


Yilmaz et al. Extended Data Fig.2

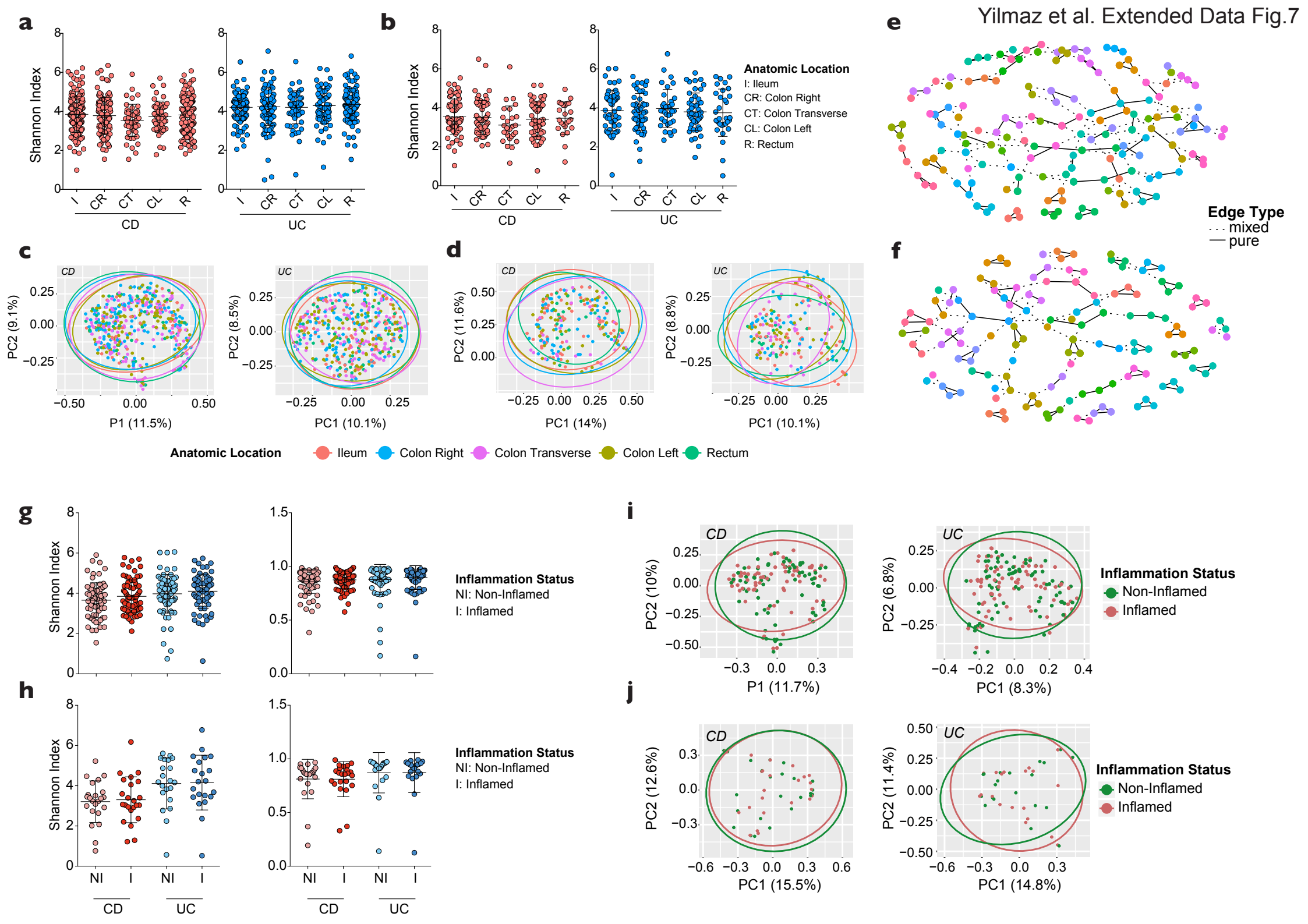

a
Cohort 1
CD
UC
Ileum
RC
TC
LC
Rectum
b
Cohort 2
Control
Actinobacteria
Bacteroidetes
Firmicutes
Fusobacteria
Proteobactaria
Spirochaetes
Tenericutes
Verrucomicrobia
c
Observed
Shannon
Simpson
Alpha Diversity Measure
p<0.05
p<0.001
d
Disease
CD
Non-IBD
UC
e
Unweighted UniFrac
Weighted UniFrac
PC1 (1.5%)
PC2 (1.1%)
PC1 (14.2%)
PC2 (12.2%)
f
PC1 (11.6%)
PC2 (7.2%)
PC1 (31.2%)
PC2 (25.2%)
g
h
Distance to centroid
CD
Control
UC
i
Higher in UC
Higher in CD
Significant Taxa
Actinobacteria
Bifidobacterium
Coriobacteriaceae
Bacteroidetes
Barnesiellaceae
Bacteroidaceae
Bacteroides
Porphyromonadaceae
Prevotella
Firmicutes
Streptococcaceae
Clostridiales
Lachnospiraceae
Ruminococcus
Blautia
Coprococcus
Lachnospira
Roseburia
Faecalibacterium
Oscillospira
Veillonellaceae
Erysipelotrichaceae
Proteobacteria
Desulfovibrionaceae
Enterobacteriaceae
Pasteurellaceae
Haemophilus
Tenericutes
Abundance Ratio (CD/UC)
j

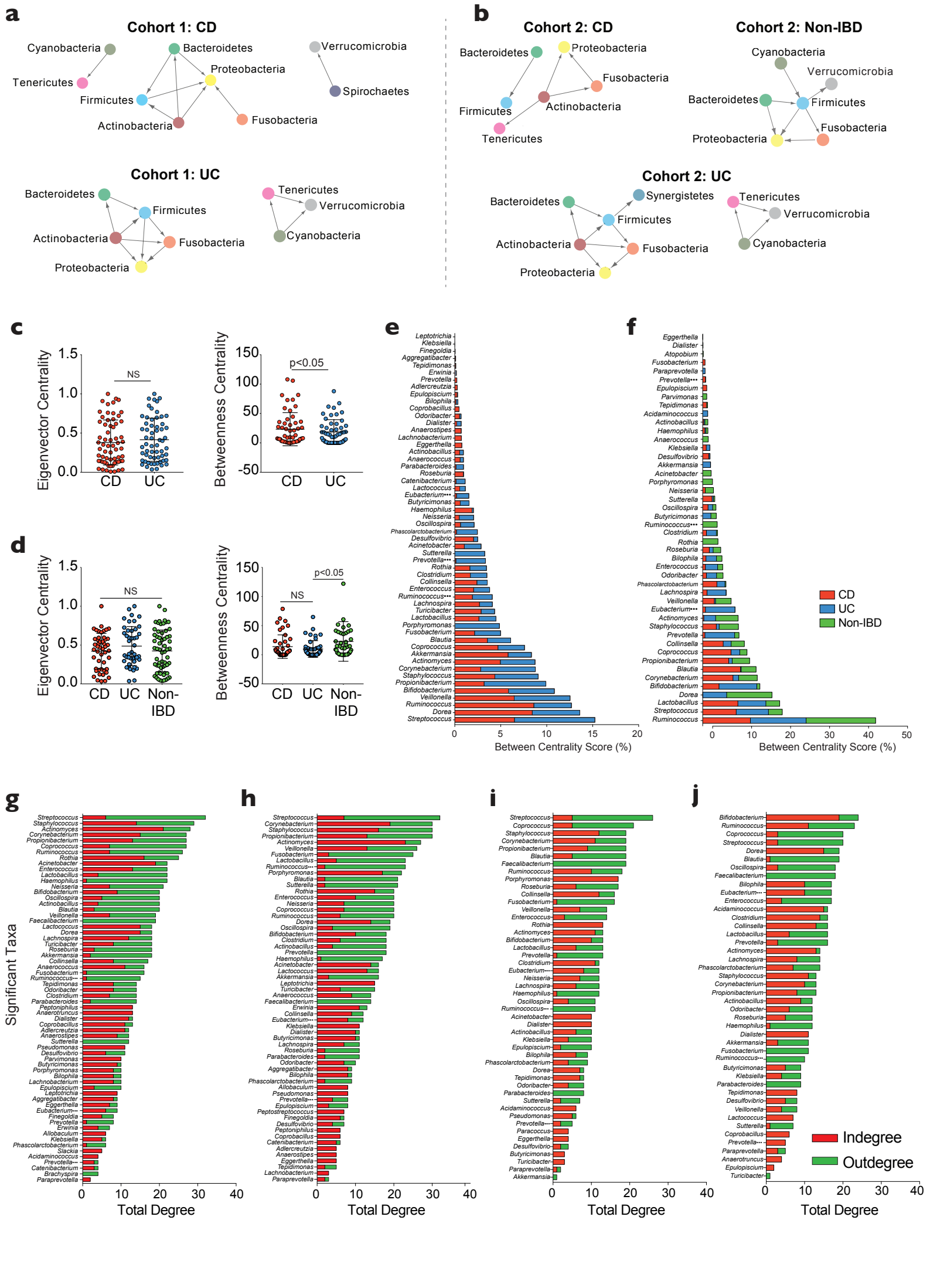

a
Higher in CD
Higher in Non-IBD
Significant Taxa
Actinobacteria
Bifidobacterium
Collinsella
Barnesiellaceae•••
Butyricimonas
Rikenellaceae•
S24-7•
Firmicutes
Clostridiales•
Lachnospiraceae•
Blautia
Coprococcus
Lachnospira
Roseburia
Ruminococcaceae•
Faecalibacterium
Ruminococcaceae••
Ruminococcus
Phascolarctobacterium
Erysipelotrichaceae•
Eubacterium•••
RF32•
Sutterella
Enterobacteriaceae•
0.5
1
2
4
8
Abundance Ratio (Non-IBD/CD)
b
Higher in UC
Higher in Non-IBD
Significant Taxa
Bifidobacterium
Collinsella
Bacteroidetes
Butyricimonas
Odoribacter
Bacteroidaceae••
Parabacteroides
Firmicutes
Clostridiaceae•
Blautia
Lachnospiraceae••
Ruminococcaceae•
Ruminococcaceae••
Ruminococcus
Erysipelotrichaceae
Proteobacteria
Tenericutes
0.5
1
2
4
8
Abundance Ratio (Non-IBD/UC)
c
denovo235444 (R.gnavus)
denovo120251 (Ruminococcaceae•)
denovo380836 (F. prausnitzii)
denovo400843 (F. prausnitzii)
denovo513014 (Ruminococcaceae•)
denovo612215 (Lachnospiraceae•)
denovo639919 (F. prausnitzii)
denovo711777 (F. prausnitzii)
denovo8857 (F. prausnitzii)
Relative Abundance (log10)
−1
−2
−3
−4
−5
CD
UC
d
denovo235444 (R.gnavus)
denovo120251 (Ruminococcaceae•)
denovo380836 (F. prausnitzii)
denovo400843 (F. prausnitzii)
denovo513014 (Ruminococcaceae•)
denovo612215 (Lachnospiraceae•)
denovo639919 (F. prausnitzii)
denovo711777 (F. prausnitzii)
denovo8857 (F. prausnitzii)
Relative Abundance (log10)
−1
−2
−3
−4
−5
CD
UC
e
100
50
0
-50
-100
PC2 (26.21%)
-200
0
200
PC1 (29.18%)
f
PC1
PC2
Coefficient
200
0
-200
p = 2.81e-06
p = 1.13e-08
UC
CD
UC
CD
g
Cohort 1
Higher in UC
Higher in CD
LPS biosynthesis
Heparan sulfate degradation
Glycerophospholipid metabolism
Folate metabolism
Fatty acid oxidation
Chondroitin sulfate degradation
Biotin metabolism
-0.1
0.0
0.1
0.2
Metabolic Subsystems Difference
h
50
0
-50
PC2 (27.15%)
-100
0
100
PC1 (36.31%)
i
PC1
PC2
Coefficient
200
100
0
-100
p = 1.22e-02
p = 3.02e-03
UC
CD
UC
CD
j
Cohort 2
Higher in UC
Higher in CD
LPS biosynthesis
Heparan sulfate degradation
Glycerophospholipid metabolism
Folate metabolism
Fatty acid oxidation
Chondroitin sulfate degradation
Biotin metabolism
-0.1
0.0
0.1
0.2
Metabolic Subsystems Difference

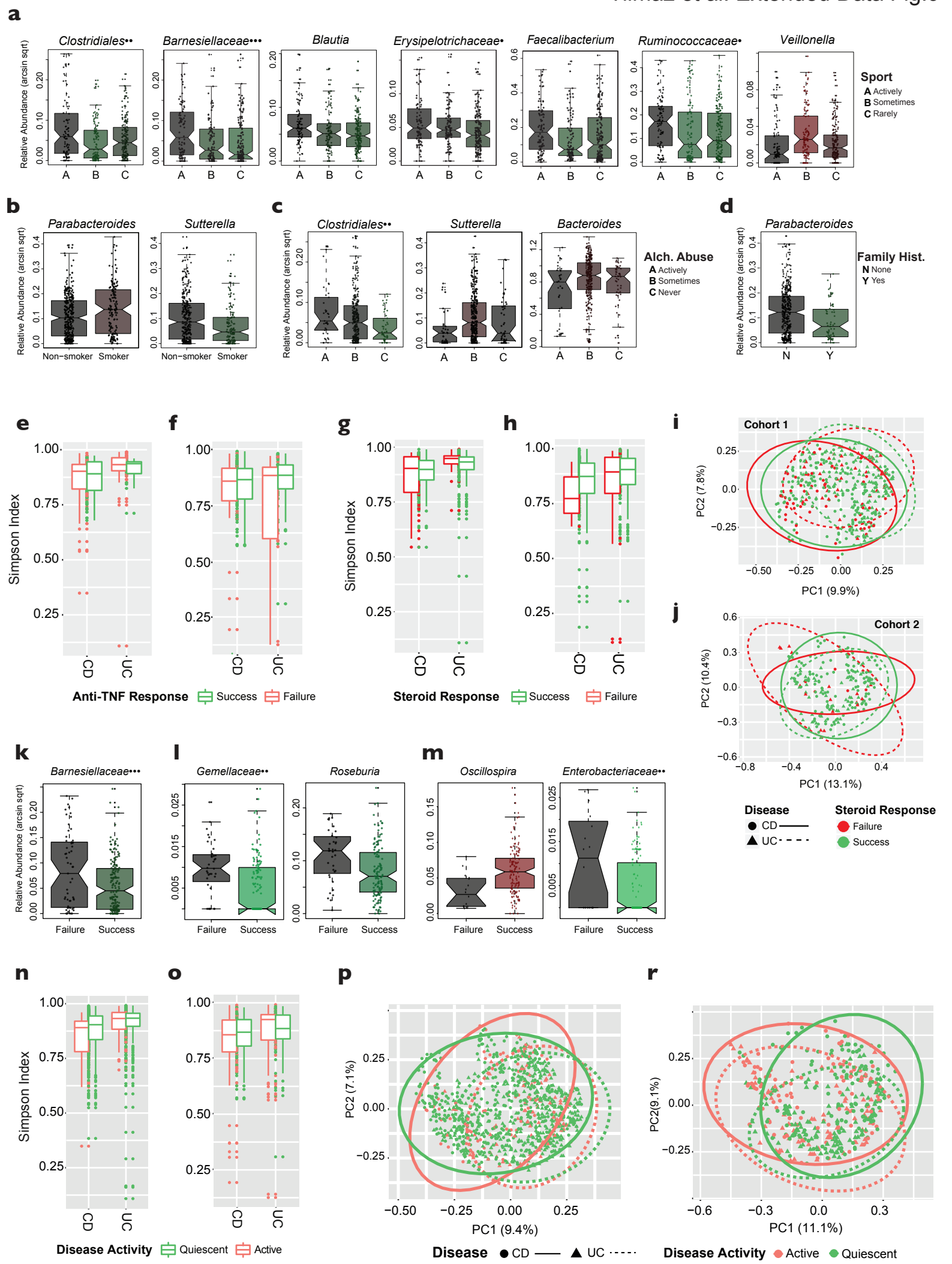

a
Model
Race/Ethnicity
Gender
Median Age
Sample Size
Species
Sample Type
Method
48 49 50 51 52 55 56 57 58 63 64 65 66 67 71 72 73 74 75 83 84 85 86 106 118 119 120 123
Median Age
Young
Adult
Old
Unknown
Sample Size
<50
50-100
101-200
>200
Species
Human
Dog
Cat
Mice
Model
Natural
DSS
Genetic+DSS
TNBS
Genetic+TNBS
Genetic
Genetic+Other
Other
Sample Type
Biopsy
Stool
Stool+Biopsy
Gender
< 50% Female
> 50% Female
Unknown
Method
FISH
RT-PCR
R454
MiSeq
HiSeq
Other
Race/Ethnicity
Mostly Caucasian
Indian
Japanese
Han
Korean
Unknown
b
Human
Dog
Cat
Mice
DSS
Genetic + DSS
TNBS
Genetic
Genetic + Other
1.0
0.5
0
-0.5
-1.0
Spearman Correlation
c
Human
Dog
Cat
Mice
DSS
Genetic + DSS
TNBS
Genetic
Genetic + Other
1.0
0.8
0.6
0.4
0.2
p<0.05
p-value (Spearman Correlation)

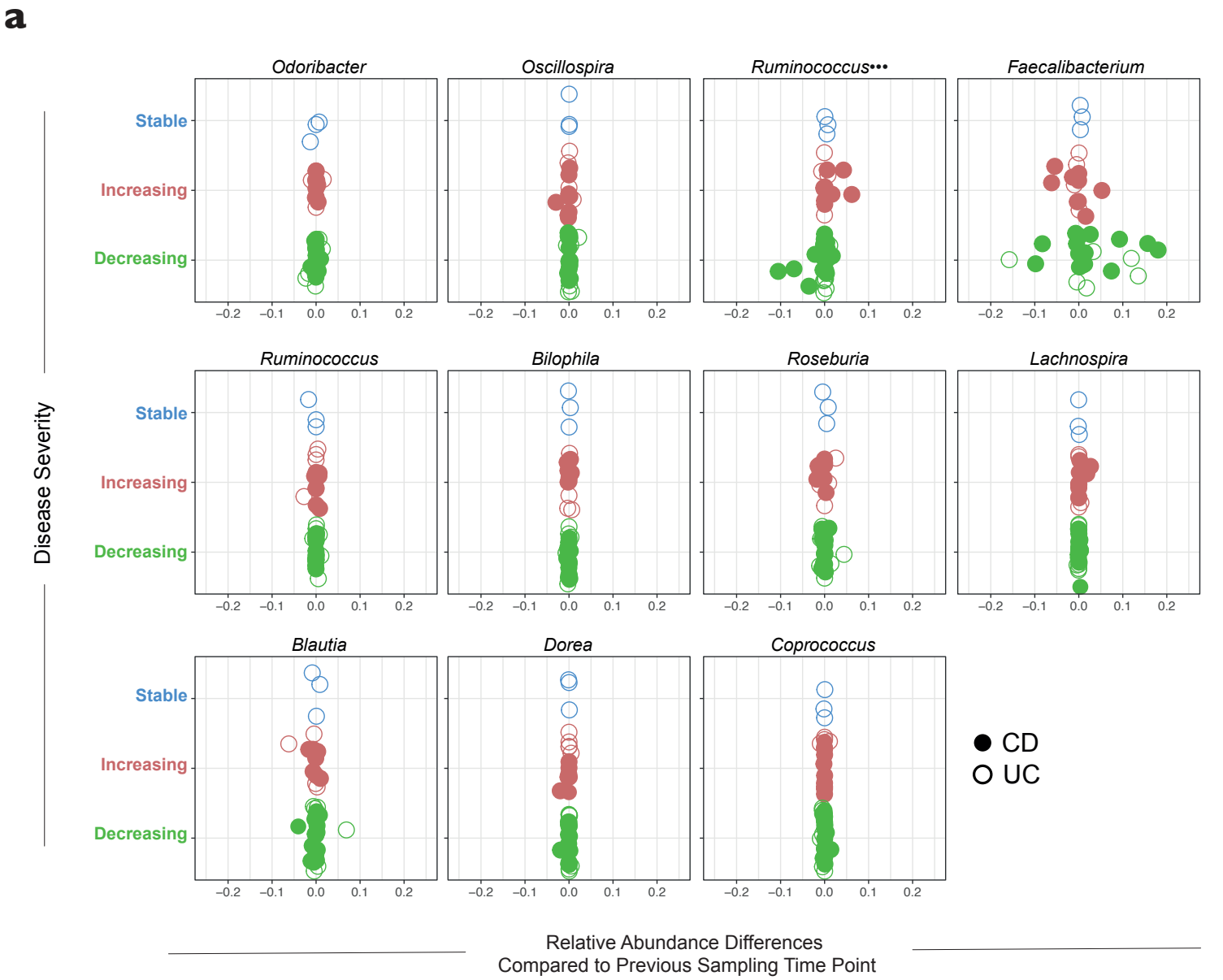

a
Patient
Cohort 1 - CD
Cohort 1 - UC
Cohort 2 - CD
Cohort 2 - UC
Years since first biopsy collection
b
Cohort 1 - CD
PC2 (13.9%)
PC1 (20.8%)
Cohort 1- UC
PC2 (16.1%)
PC1 (28%)
Cohort 2 - CD
PC2 (12.5)
PC1(18.8%)
Cohort 2 - UC
PC2 (22.4%)
PC1 (29.6%)
c
Bacteroidetes
Firmicutes
Proteobacteria
Disease Severity
Stable
Increasing
Decreasing
CD
UC
Relative Abundance Differences
Compared to Previous Sampling Time Point
d
CD
Enterobacteriaceae•
Klebsiella
Relative Abundacnce
(arcsin sqrt)
Fecal calprotectin (µg/ml)
e
UC
Ruminococcus•••
Prevotella
Relative Abundacnce
(arcsin sqrt)
Fecal calprotectin (µg/ml)

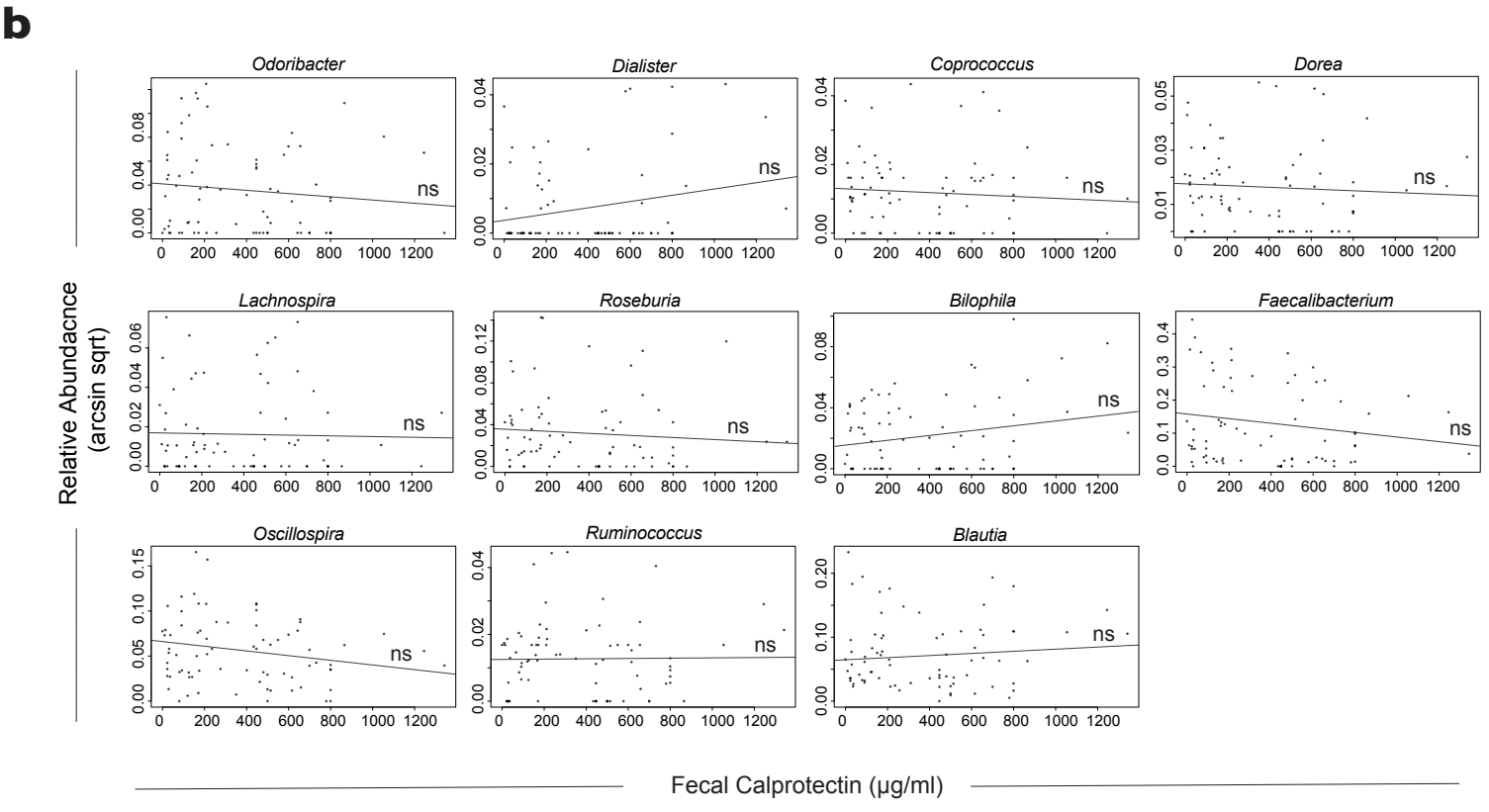
Yilmaz et al. Extended Data Fig.7




a
Cohort 1: CD
Cyanobacteria
Bacteroidetes
Verrucomicrobia
Proteobacteria
Tenericutes
Firmicutes
Spirochaetes
Actinobacteria
Fusobacteria
Cohort 1: UC
Bacteroidetes
Tenericutes
Firmicutes
Verrucomicrobia
Actinobacteria
Fusobacteria
Cyanobacteria
Proteobacteria
b
Cohort 2: CD
Bacteroidetes
Proteobacteria
Fusobacteria
Firmicutes
Actinobacteria
Tenericutes
Cohort 2: Non-IBD
Cyanobacteria
Verrucomicrobia
Bacteroidetes
Firmicutes
Fusobacteria
Proteobacteria
Cohort 2: UC
Bacteroidetes
Synergistetes
Tenericutes
Firmicutes
Verrucomicrobia
Actinobacteria
Fusobacteria
Cyanobacteria
Proteobacteria
c
Eigenvector Centrality
NS
CD
UC
Betweenness Centrality
p<0.05
CD
UC
d
Eigenvector Centrality
NS
CD
UC
Non-IBD
Betweenness Centrality
p<0.05
NS
CD
UC
Non-IBD
e
Between Centrality Score (%)
f
CD
UC
Non-IBD
Between Centrality Score (%)
g
Significant Taxa
Total Degree
h
Total Degree
i
Total Degree
j
Indegree
Outdegree
Total Degree


a
Clostridiales••
Barnesiellaceae•••
Blautia
Erysipelotrichaceae•
Faecalibacterium
Ruminococcaceae•
Veillonella
Relative Abundance (arcsin sqrt)
A B C
Sport
A Actively
B Sometimes
C Rarely
b
Parabacteroides
Sutterella
Non−smoker Smoker
c
Clostridiales••
Sutterella
Bacteroides
Alch. Abuse
A Actively
B Sometimes
C Never
d
Parabacteroides
N Y
Family Hist.
N None
Y Yes
e
f
g
h
Simpson Index
CD UC
Anti-TNF Response Success Failure
Steroid Response Success Failure
i
Cohort 1
PC2 (7.8%)
PC1 (9.9%)
j
Cohort 2
PC2 (10.4%)
PC1 (13.1%)
Disease
CD
UC
Steroid Response
Failure
Success
k
Barnesiellaceae•••
l
Gemellaceae••
Roseburia
m
Oscillospira
Enterobacteriaceae••
Failure Success
n
o
Disease Activity Quiescent Active
p
PC2 (7.1%)
PC1 (9.4%)
Disease CD UC
r
PC2(9.1%)
PC1 (11.1%)
Disease Activity Active Quiescent


a
Odoribacter
Oscillospira
Ruminococcus•••
Faecalibacterium
Ruminococcus
Bilophila
Roseburia
Lachnospira
Blautia
Dorea
Coprococcus
Stable
Increasing
Decreasing
Disease Severity
CD
UC
Relative Abundance Differences
Compared to Previous Sampling Time Point


b
Odoribacter
Dialister
Coprococcus
Dorea
Lachnospira
Roseburia
Bilophila
Faecalibacterium
Oscillospira
Ruminococcus
Blautia
ns
Relative Abundacnce
(arcsin sqrt)
Fecal Calprotectin (µg/ml)